

Beyond the trouble in the Gulf

No amount of resolution by the United Nations can hide the truth that even a quick resolution of the problems created in Kuwait by Iraq (which seems unlikely) will not restore stability to the Middle East.

THE general consternation that the end of the 40-year Cold War should so swiftly be followed by the prospect of serious trouble in the Middle East is understandable, but this disconcerting turn of events is not surprising. Iraq, which annexed Kuwait at the beginning of the month, has only just reached an uneasy stalemate with its neighbour Iran after eight years of warfare that left its economy in ruins and few of its military objectives attained. The general indifference of governments elsewhere to that conflict, both middle-sized powers in the West (such as Britain and France) as well as the then two superpowers (the Soviet Union and the United States) is chiefly explicable by the way in which both combatants kept supplies of oil flowing. Each, after all, had to pay for its military equipment. Such interest as the Iran-Iraq war evoked resulted in mild support for Iraq, given the general distaste for the previous Khomeini regime in Iran. But now the shoe is on the other foot.

Iraq makes no secret of its awkwardness as an adversary. It is equipped with chemical weapons (poison gas) and has experience in their use, both in the war with Iran and against its own population of Kurds, in northern Iraq. It has a supply of long-range surface-to-surface rockets, but little experience of their use. Its flirtation with means for making a huge gun was fanciful, but has probably come to nothing. Iraq is not yet a nuclear power, and is unlikely ever to be one in its own right; one of the disturbing features of this month's developments is that the combination of Iraqi determination and Kuwaiti money may persuade some other states to join in such a venture. But the most menacing development of all is Iraq's apparent willingness to use foreigners caught by accident in its extended borders as hostages in future dealings with governments elsewhere. The unanimity of the UN Security Council in the days following the annexation of Kuwait remains the best hope that order will be restored in the Middle East by legal processes, not military conflict. But it remains to be seen whether the resolution of governments elsewhere will survive televised appeals for compromise by their trapped nationals.

That is one reason why attention should be given now to the causes, rather than the occasion, of the conflict in the Middle East. The underlying difficulty is the revenue from the sale of crude oil earned by states in the Middle East, and the fortuitously unequal distribution of that income. Iraq, with 15 per cent of the total production of

the Organization of Oil Exporting Countries (OPEC), was not badly off, but its income from the sale of \$20-a-barrel oil was not sufficient to pay for the war with Iran and is plainly insufficient to sustain its plans to be a substantial military force among Arab states (whence, of course, the move on Kuwait). But Iraq is at least distinguished from the other oil-producers of the Middle East by its plan for spending its oil revenues on a national programme of some kind, being militarily strong in this case. The bulk of what other oil-producing states earn is routinely lent back to the purchasers of oil, usually in the form of credits controlled by named individuals rather than by OPEC governments. The chief effect of 20 years of high-priced oil is that it has created many millionaires among the populations of the Middle East.

The conspicuous failure of the same 20 years has been the poverty of constructive development in the oil-producing states of the Middle East. It is true, of course, that the wealthier governments have set out to provide their people with advanced health services, usually relying on physicians and other paramedical workers from elsewhere to service them. There have also been developments in education, but they are mostly unremarkable. The oil-producers of the Middle East, blessed though they have been with income of the order of \$200,000 million a year from oil, seem not to have turned their minds to creating a first-rate university somewhere. (The University of Cairo, in impoverished Egypt, remains the Arab world's outstanding university.) Similarly, there has been too little investment in industrial enterprises, even in the downstream processing of oil itself. The oil producers have also spent substantial parts of their funds on preparation to defend themselves and their fortunes; the experience of Kuwait shows how effective that has been. It is no wonder that Kuwait has fallen to Iraq.

There is no way in which the governments now joined in the enterprise of forcing Iraq to retreat from Kuwait can force more enlightened policies on the undemocratic governments elsewhere in the region, but they can at least say loudly and publicly what they think is wrong. Both the vulnerability of the Arab oil-producing states to attacks such as that of Iraq on Kuwait and to Shiite rebellion and revolution are simply consequences of their hierarchical social structure and nepotistic ways of sharing unimagined wealth. While these practices continue, troubles like that which has now arisen are bound to recur. □



Integration problems

Science has a fighting chance of survival in the eastern part of Germany. But will it seize the moment?

WHAT is to be done to integrate the research establishment of East Germany, hitherto supported from Berlin, with that of West Germany, supported from Bonn? It has been plain since the outset of the reunification project that there would be serious difficulties in absorbing East Germany's stalinist Academy of Sciences into the quite different pattern of research supported by, among others, the Deutsche Forschungsgemeinschaft (DFG) and the Max-Planck Gesellschaft (MPG), many of which were vividly made plain by the letter from Dr Werner Franke last week (*Nature* 346, 603; 1990). Put bluntly, the problem in East Germany is that almost all research is supported through the academy and carried out by institutes whose tenured members conduct research which is often second-rate and whose personal reputations have been sullied by their association with the regime whose last vestiges will disappear with the elections planned for December. What will happen to the eastern academy institutes after that?

Wisely, the West German research minister, Heinz Riesenhuber, has agreed that 18 months should pass before the fate of the institutes and their researchers is finally decided. To have acted otherwise would have been to emulate old-fashioned East German heavy-handedness. The hope is that the interval will allow a sober evaluation of the parts of the old academy system worth saving. But the process is bound to leave scars.

So how should this breathing-space be spent? One possibility is that the academy system may be able to shrink itself to a more modest size. That is what Riesenhuber himself expects. The snag is that the end result would be a series of institutes not necessarily of a kind that could be integrated into the West German system. It would be much better that the interval should be spent in allowing elements of the research enterprise in the east, from individuals to whole institutes, who consider they have a legitimate claim on funds from the West should use the interval to apply for support from DFG, MPG or other western sources. That would also be a painful process, but one that people could understand.

However the integration is accomplished, it will not be free from recrimination. Under the old regime, the patronage of the stalinist government was a necessary requirement for even modest success in science. People at odds with the ideology were unable to meet their peers by travelling abroad, and were often denied facilities for research. To what extent, now, should there be positive discrimination in favour of these people? And may not those who sold their soul to the old regime also be seen as victims? That is a question that will keep the evaluators awake at nights, not only in East Germany but in the rest of reforming Eastern Europe. □

Untransparent grants

The British Medical Research Council will have to be more open about its research policy.

DURING the past ten hard years, the British research councils have miraculously managed to keep the respect of their constituents, researchers at British universities, by the transparent fairness of their decisions. Even when there has not been enough money to go around, grant applications have been scored with apparent objectivity by the peer-review system, and many applicants destined for disappointment have been told that their proposals would in normal times have been supported. But there are now signs that the strains on the system may be undermining it. The fuss about the refusal of the Medical Research Council (MRC) to renew a research grant at Cambridge (see opposite) is a sign that strains on the British research-support system are now insupportable.

The circumstances are not entirely clear, although it seems to have been determined that the ten-strong Medical Cryobiology Group (which costs £250,000 a year) is to be disbanded. The research bears on the physics and the physiology of the preservation of organs for transplantation in which Cambridge is a leading British centre. Referees (not exclusively from Britain) appear to have given the work at Cambridge glowing testimonials, but the MRC Cell Board seems to have offset these high opinions with the conclusion that the research, while interesting, lacks urgency and, indeed, the potential for making a mark on research internationally. At least some members of the subcommittee that reached that conclusion are now dissenting from the decision to close down the research group, saying that their opinions have been misrepresented. Transplant surgeons, in Britain and elsewhere, are meanwhile up in arms.

The issue of principle raised by these developments concerns the degree to which grant-making bodies such as the MRC must make their decisions intelligible to the wider world. Especially when research funds are as scarce as they have become in Britain, nobody expects that all decisions can be clear-cut, or determined exclusively by scientific quality. Questions of whether the research concerned can be afforded a place in some national pattern of activity are also legitimate and proper. What the MRC appears on this occasion not to have appreciated is that decisions based on these broader considerations must also be argued publicly if they are to make sense to those directly affected and to the research community in general. It will be disgraceful if the MRC has sought to avoid the difficulty of justifying an unwelcome decision by concealing its reasons for reaching it, which is the obvious inference of the Cell Board's report. In any case, the MRC has no choice but to argue the case for strategic decisions publicly. Or rather, if it chooses to act otherwise, it must recognize that it will lose the support of its research community. □

MRC's cryobiology closure evokes protests

- Reviewers' recommendations over-ruled
- Lack of clinical results cited

London

THE United Kingdom's Medical Research Council faced a storm of protest from British and international research communities last week over its decision to close a small cryobiology unit in Cambridge, the only group developing techniques for freezing human tissue and organs in Britain.

Condemned by leading biomedical researchers on both sides of the Atlantic and as far away as Australia, the decision will lead to the dispersal an internationally valued team of 10 researchers. All told, the group, led by David Pegg, currently receives £250,000 a year out of the MRC's budget of £180 million. At the heart of the furore over the closure is the accusation that the MRC wrongly overruled the opinion of a panel of expert referees who were unanimous in recommending continued funding of the unit. One of them, Professor Felix Franks of the University of Nottingham, has even denounced the MRC decision as "a corruption of the peer review system".

Appointed last year by the MRC to conduct a routine review of the Cryobiology Unit's work, the referees all submitted highly favourable reports. But earlier this year the council's Cell Board — one of several committees that are responsible for making funding decisions — acting on a report from its own subcommittee, gave the unit only a 'beta' rating, a verdict that, with many 'alpha'-rated projects currently failing to win funds, guaranteed the unit's closure.

The Cell Board judged the unit's work to fall below the "competitive standard" in the current financial climate, says Nick Winterton, head of the secretariat of the MRC. That judgement, while apparently at odds with the views of specialists, reflects the board's "wider scientific representation", he adds. Effectively, the unit's fate was sealed by a concluding remark in the subcommittee report that, despite 10 years of funding, clinically useful applications seem to be "still around the corner".

Be that as it may, the most outspoken critics of the decision are transplant surgeons and clinicians. In a letter of protest, Sir Roy Calne, Professor of Surgery at Cambridge, said he was "amazed" at the decision and that he "earnestly requested the council to reconsider". The MRC, says Calne, does not appreciate the enormity of the problem being tackled by the group — how to keep tis-

sue at sub-zero temperatures without it being damaged by ice crystals — which is "probably as difficult a thing to crack as the immune system in organ rejection". At least 70 other clinicians and scientists have also lodged protests.

These might have proved easier for the MRC to deflect if the two cryobiologists invited to sit on the Cell Board's subcommittee had not already voiced dissent. One of them, Peter Mazur of Oak Ridge National Laboratory in Tennessee, complained to the Cell Board's chairman saying that his positive views had been ignored and describing the report's emphasis on "clinical isolation" as "a distortion". In a letter to Dai Rees, head of the MRC, Mazur said that he had been "thunderstruck" to hear of the recommendation to close the unit. According to

DRUG APPROVAL

Fast track for AIDS and cancer drugs

Washington

NEW drugs for life-threatening diseases such as cancer and AIDS should be handled more quickly by the US Food and Drug Administration (FDA), concludes a report released last week by the President's Cancer Panel. The recommendations come from a committee, chaired by Louis Lasagna of Tufts University, asked by the Cancer Panel to determine whether the approval criteria for cancer or AIDS drugs should be modified, and to identify barriers that have impeded the access of cancer and AIDS patients to new treatments.

The committee concurs with a recent Government Accounting Office opinion that the FDA is under-financed and undermanned. But a substantial increase in federal funds for FDA is unlikely in the present fiscal climate, and the committee recommends that the FDA should reduce its burden by relying more heavily on outside reviewers. Phase I and II investigational new drug (IND) applications that are filed by academic "non-commercial" researchers, many of which involve new uses for approved drugs, should be reviewed by Institutional Review Boards, the committee says. The report also suggests that drug companies should be given the option of paying "user fees" to the FDA for an expedited review of a drug by experts outside the FDA who have no conflict of interest.

To hasten the approval process for cancer and AIDS drugs, the committee

Winterton, the original report was not amended but the reservations of the two dissenting cryobiologists were communicated to the Cell Board before it made its final recommendation.

For some critics, news that a medical transfusion department currently being set up in Cambridge University had intended to incorporate the cryobiology unit as its main research arm made the council's decision even harder to swallow. Such a move would have answered the chief criticism of clinical isolation, says Calne. And concern over the closure rose further still when it emerged that plans to establish a new bank for human heart valves in Cambridge had been shelved because of the MRC's decision.

The planned closure is assumed by many critics to reflect on the need for a strategic reorganization of funding within the MRC rather than the quality of the unit's work. But regardless of any protest, it is unlikely that the MRC's decision will be reversed — the Cell Board has already had ample opportunity to review its original decision in the light of representation, says Winterton.

David Concar

urges the FDA to make its criteria more flexible: for example, in assessing the therapeutic benefit of a drug for the treatment of life-threatening condition, proof that the drug will prolong life need not be a prerequisite for FDA approval. Instead, The report suggests that a drug that produces tumour regression in more than 20 to 30 percent of patients who have not responded to alternative therapies should be approved.

Similarly, the development of AIDS drugs could be facilitated by the approval of drugs that not only cause a rise in the T4 cell count but also improve a patient's quality of life.

The committee also believes that phase III cancer studies, designed to compare the efficacy of an experimental drug with that of a marketed drug, can delay the drug approval process and should not be a requirement for marketing approval.

The committee supports the rights of patients with life-threatening diseases to obtain accelerated access to experimental drugs, and although Lasagna accepts that making new drugs available for marketing at an earlier stage may increase the dangers of toxicity, he believes that "more good than harm" would come from implementing the committee's recommendations. The committee points out that patients with life-threatening conditions, for whom there are no alternative treatments, are often willing to accept the greater risks associated with new therapies.

Diane Gershon

Radio to the rescue

Washington

SIMILARITY between the distorted optical images produced by the out-of-focus Hubble Space Telescope (HST) and the signals normally generated by multi-element radioastronomy telescopes will allow some parts of the HST's planned observing programme to be salvaged, astronomers said last week. Image processing techniques developed over the past decade by the National Radio Astronomical Observatory (NRAO) have proved surprisingly effective in filtering out the blurry 'haloes' that are the principal symptom of the HST's malaise.

Rick White, of the Space Telescope Science Institute in Baltimore, says that after HST's problem was announced, "we noticed right away that the fuzzy images, with a well-defined core and nasty side lobes, looked just like radioastronomy data". The HST images, produced by an optical system whose primary mirror suffers from spherical aberration, have a sharp central core surrounded by a diffuse halo of improperly focused light (see *Nature* 346, 3; 1990). This image structure is characteristic of intensity maps produced by radio interferometers, in which signals from separate radiotelescopes are combined to produce an image whose resolution is much higher than that of the individual instruments.

Because the aberration in the HST's mirror is symmetrical and well characterized, the halo of unfocused light around each stellar image is very predictable and can be mathematically subtracted, says White. Although the size and shape of the halo are somewhat frequency-dependent, filters can be used to produce monochromatic images suitable for computer processing.

Photographs released last week by the National Aeronautics and Space Administration (NASA) show how image processing resolves individual stars within the star cluster R136, part of the 30 Doradus nebula (see below). The upper image was made with the Wide Field/Planetary Camera on 3 August; the unprocessed enlargement at lower right shows about 60 stars, more than can be seen from the ground, but in the image at lower left, enhanced using computer codes derived from the NRAO algorithms, hundreds of stars stand out. Beyond its technical significance, this image represents a genuine scientific achievement — the measured density of this cluster of young, massive stars is higher than anything previously seen.

Although image processing will restore much of the HST's full resolution, according to White, the dynamic range (the ability to detect faint objects when they are close to bright ones) cannot be so easily restored. In HST's blurred images, only 15 per cent of the total light is in a central core, rather than the 70 per cent that a perfect mirror would have provided. The rest of the light spills over into the halo, swamping any nearby faint sources. What that means, says White, is "that there's a fundamental limit to how much you can improve the image. No image-processing technique is going to put something in the data that isn't there". Similarly, the reduction in the brightness of the central core means that the HST is in absolute terms less sensitive than it should have been. Until it is fitted with a corrective lens in perhaps three years' time, the telescope will be about a factor of 10 less farsighted; the very faint galaxies and quasars that were to have been the

HST's most distant goal will remain hidden until then. Nevertheless, if observation periods are extended and the new image-processing techniques used, White thinks that as much as 80 per cent of the science planned for the HST's first two years appears to be back within reach.

Christopher Anderson

■ A statement released last week by Lew Allen, chairman of the HST Board of Investigation, confirmed that the entire half-wavelength error in the primary mirror's shape could be attributed to a 1.3-mm positional error in one of the components of the reflective null corrector (see *Nature* 346, 599; 1990).

Tantalizingly, Allen observes that some of the spacing washers in the null corrector have a thickness of 50 thousandths of an inch — or 1.3 mm. A misplaced washer may turn out to be the ten-cent cause of a billion-dollar mistake. **D.L.**

LARGEST TELESCOPE

Go-ahead for Japanese telescope?

Tokyo

JAPAN'S Ministry of Education, Science and Culture (MESC) has at last decided to start building the world's largest telescope, according to a leak in the Japanese press on Monday.

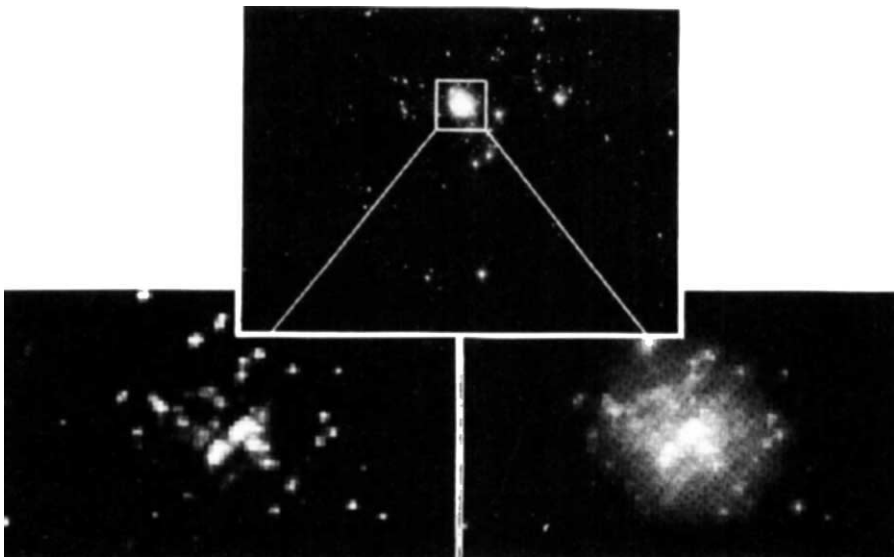
Plans for the 7.5-metre single-mirror telescope, to be built in Hawaii, have been in the works for years (see *Nature* 311, 5; 1984), and this fiscal year MESC acquired 13 million yen (\$90,000) in 'preparatory' funds for the project. But MESC officials have insisted that a final go-ahead for the project, which is eventually expected to cost 45,000 million yen (\$300 million), has not yet been given (*Nature* 343, 295; 1990).

According to a report in the Yomiuri newspaper on Monday, MESC will apply at the end of this month for 600 million yen (\$4 million) to begin construction of the telescope in fiscal 1991.

Keiichi Kodaira of the National Astronomical Observatory, a well known promoter of the project, could not confirm the Yomiuri report on Monday. He says a final decision is expected on August 22 or 23 and will be officially announced at the beginning of September. But he says he has the impression MESC is "very positive".

The telescope has competition within MESC from another large project in the form of Superkamiokande, a proposal from Tokyo University for the world's largest observatory for neutrinos and proton decay (*Nature* 331, 379; 1988). Supporters of that project are also fairly confident of winning MESC funds to commence construction next year, but they too are waiting for an official announcement at the end of this month.

David Swinbanks



Radioastronomy techniques succeeded in teasing out the information from Hubble's original effort. The merits of a partially functioning Hubble are the subject of a letter in the Correspondence pages.

Ready for the future?

Pullman, Washington

LAST week, 3,000 agricultural scientists around the world saw the future, and apart from a few technical glitches, dead spots and occasional lack of focus, it worked. An international symposium on plant biotechnology put on by Washington State University was the first interactive televised scientific conference to be broadcast worldwide, and although it was not without hitches it showed that that satellite networks are finally up to the task of video conferencing.

The symposium, a year in planning, illustrated strengths and weaknesses of teleconferencing technology. With an audience distributed at 70 viewing centres worldwide, it brought together plant biotechnology experts with scientists in developing countries who might not have been able to attend an international meeting. Conference officials estimate that at about half of the viewing centres, local organizers used the three-hour symposium as the nucleus of longer local meeting. Observers at all sites were able to fax written questions during the conference to Pullman; by mid-point, over a hundred questions had been received. Several dozen questions were answered on-air, and the others received written responses.

The conference featured two panels of plant biotechnologists, one group in Pullman, the other at the Max-Planck Institute for Genetics Research in Cologne, West Germany. Pullman's five-member panel was composed of scientists from the United States, Mexico and Australia, along with an American journalist — Jack Reynolds of the United States Information Agency — who served as moderator. In Cologne, Jozef St. Schell, director of the Max-Planck Institute, led a four-person panel that included former International Rice Institute director M. S. Swaminathan and Konstantin Skyrabin of Moscow State University.

Except for a brief loss of audio from Cologne and some minor production glitches, the conference was technically flawless.

Much of the success was due to redundancy in the form of duplicate equipment at every stage of link (which included seven satellites and an equal number of ground stations), and a large production crew of professionals and students from Washington State University's telecommunications programme.

But as technically impressive as the symposium undeniably was, it illustrated some of the traditional problems of international conferences, as well as some new ones. With an audience of biotechnology researchers, Third World agricultural scientists, students, and even Washington

State legislators, the conference oscillated wildly in tone, technical content, and direction.

A basic 15-minute taped primer on genetic engineering by Schell was immediately followed by a debate on the fine points of RFLP (no definition given) techniques by other panel members. Summaries given at the end of each one-hour segment appeared to have been largely prepared beforehand, reflecting little of the unscripted debate. And a prepared presentation on the application of biotechnology to world food production by Robert Fraley of the Monsanto Corporation (one of the conference sponsors) came dangerously close to becoming a video advertisement for Monsanto products.

The basic problem, however, appeared to be that because little more than a 5-metre satellite dish and a television set was needed to plug in to the conference, the audience became impossibly broad, ranging from leading researchers to interested but nonscientific members of the public. The organizers tried to include a little something for everyone, with the result that few participants seemed to find the conference very satisfying. An exit survey taken at the Pullman site showed that the symposium did a poor job of demonstrating research techniques, conveying useful scientific information or providing new research ideas.

On the other hand, most participants gave the conference high marks for providing access to scientists who would not otherwise have been available. And many said that the conference had been their first introduction to the concerns of the Third World — the issues of access to research, costs, and preservation of traditional farming techniques. In one exchange that clearly illustrated the debate within the developing world on the hazards of biotechnology, Luis Herrera-Estrella of the Mexican Centre for Research and Advanced Studies (and a member of the Pullman panel) said that Mexico "is most interested in food production. The ecological impact [of biotechnology] is less important". But Indian scientist Swaminathan, speaking from Cologne, quickly jumped in to warn against the use of the Third World as a biotechnology laboratory. "I disagree that it does not matter as much in the developing world. Ecological groundrules must be the same wherever they are on the planet", he said.

Although the jury is still out on the scientific success of teleconferences, the symposium showed that electronic meetings can be substantially cheaper than their traditional equivalents. Total costs amounted to about \$300,000, the majority

Trial approved for saliva test

San Francisco

A NOVEL saliva-based test for HIV (human immunodeficiency virus) infection, which may ultimately lead to a 'do-it-yourself' AIDS test, has been approved by the US Food and Drug Administration (FDA) for human clinical trials. The technique, developed by Epitope Inc. of Beaverton, Oregon, may make AIDS testing simpler and more widely available because antibodies are collected from the mouth without the need to draw a blood sample.

In the test, a treated pad is placed between the gums and cheek for two minutes to collect antibodies to HIV, which are present in the saliva. The pad is then placed in a vial and mailed to a laboratory where it is submitted to the standard HIV test now used on blood samples. In pre-clinical trials on 600 patients, Epitope scientists report that the collection device has been perfectly reliable, leading them to believe the test will be "fully equivalent" to blood tests.

FDA approval will allow Epitope to carry out clinical trials for nine months at five sites nationwide, where matched samples of blood and saliva will be compared. If marketing approval is then granted, the company plans to make the device available worldwide. Potential users include police and emergency medical personnel who request frequent testing for HIV, and developing countries where blood testing may be impractical.

The company is now considering ways to address questions of counselling and anonymity that would arise with widespread use of the test. One option under consideration is that AIDS hotlines or counselling centres would be contracted to inform users of the results of their tests, and of what can be done following a positive result. Epitope hopes ultimately to produce an AIDS test, based on the saliva collection device, which could be conducted entirely at home.

Elizabeth Schaefer

of which was covered by in-kind donations from a half-dozen telecommunication companies. A traditional conference with the same attendance would typically cost the organizers about \$1 million and the participants another \$3 million for travel and lodging, according to Washington State University administrator Donald Hanna, who co-chaired the event.

But the larger lesson, he said, is that successful global teleconferences will have to cover fields that are changing too fast for the scientific journals to keep up with, stick to subjects like plant biotechnology and involve issues of worldwide importance, such as health, food, or population.

Christopher Anderson

Fallout hits research

Munich

OPPOSITION to nuclear power, a cause embraced by citizens' initiatives and splinter parties a decade ago, has begun to take over the political mainstream in West Germany, with potentially dire consequences for scientific research. In a decision that is certain to have broader implications, West Berlin Mayor Walter Momper, a Social Democrat, last week surprisingly withdrew his support for a 10-MW research reactor located at the Hahn-Meitner Institute (HMI). The move endangers the existence of HMI, one of 13 national laboratories in West Germany.

Momper had originally promised to award a licence to the refurbished reactor by 31 July at the latest. But the environment minister in his coalition government, Michael Schreyer of the *Alternative Liste*, refused to issue the licence because of the lack of permanent storage space for the radioactive waste produced by the reactor (see *Nature* 346, 405; 1990.) *Alternative Liste* is the local branch of the environmentalist Green Party.

After a majority vote on 7 August by the West Berlin Senate in favour of licensing the reactor, researchers had been expecting Momper to reverse Schreyer's decision somehow, for example by removing her jurisdiction over licensing. But such a move would have endangered the fragile West Berlin coalition, and elections are only three months away. Instead, Momper returned from holidays on 13 August with the surprising news that he had been at least partly convinced by Schreyer's arguments.

Momper then offered HMI two unappetizing options. If he found Schreyer's case against licensing the reactor "watertight", he said, there would be nothing more he could do, and it would be up to HMI to bring a lawsuit against Schreyer. If not, he would reopen the licensing procedure, which would probably require new public hearings and could not be concluded until after the forthcoming elections. In either case, it would take months, if not years, for the reactor to receive a licence.

The overhaul and upgrading of the reactor from 5 to 10 MW was completed last year at a cost of DM170 million (about \$110 million). Most of the money came from the Research Ministry in Bonn, which pays 90 per cent of the operating costs of the national laboratories (*Grossforschungsanstalten*). The reactor is to be used as a neutron source for biology, medicine and chemistry as well as physics.

Tired of broken promises from Momper, West German Research Minister Heinz Riesenhuber (Christian Democrat) last week issued an ultimatum: license the reactor by 21 August, or Bonn

will pull out of financing HMI. According to West Berlin Senate spokesman Werner Kolhoff, West Berlin is already overburdened by the costs of reunification and could not afford the DM100 million a year that would be required to keep HMI open if Bonn pulls out.

Some see Momper's and Riesenhuber's actions as gestures made solely to avoid losing face. As a prominent representative of a party that is officially against nuclear energy, Momper cannot afford to back the HMI reactor unless final waste storage is secured. And Riesenhuber, who has been threatening to pull out of HMI for months, cannot easily back down now.

On this view, HMI will have to be patient for a few months until all-German and all-Berlin elections, both of which are scheduled for 2 December. After that, perhaps, the federal government will be able to force the licensing of the reactor in the case of a legal challenge, a possibility that comes about because of the murky legal status of West Berlin. For at least a few more weeks, it will remain under the jurisdiction of Allied occupying forces, not of Bonn.

But the patience of HMI is nearly at an end. Spokesman Thomas Robertson called the situation at HMI "miserable". HMI has lost one candidate for the open position of director and may well lose a second in the next few days. Researchers in unrelated fields such as solar cell development who are contemplating moving to HMI are bound to be influenced by the worries about the institute's future, says Robertson.

The turmoil at HMI occurs against a backdrop of rising opposition to nuclear power in West Germany. The Social Democrats (SPD), one of the two largest political parties, dropped their support of nuclear power in 1987 following the Chernobyl accident. SPD has won a series of elections in the *Länder* (states) which have given it a majority in the Bundesrat, the upper house of parliament. A victory by SPD in forthcoming national elections could turn the West German nuclear industry on its head, with an equally devastating if narrower impact on science. The problem may become especially acute at Garching, near Munich, where the Bavarian government has hesitated to replace an ageing research reactor at the Technical University of Munich because it feared a local backlash.

The dispute about the HMI reactor has already endangered Berlin's standing as an attractive site for research and industry. Both the heads of West German research organizations and the Chamber of Commerce and Industry of West Berlin have issued messages to that effect in recent weeks.

Steven Dickman

MITI heads for inner space

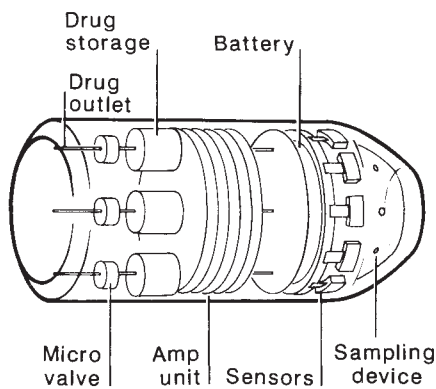
Tokyo

JAPAN'S ever-ambitious Ministry of International Trade and Industry (MITI) seems to have found in Steven Spielberg's film *Inner Space*, about a microsubmarine that ventures through human blood vessels, an idea for its next venture. At the end of this month, MITI's Agency of Industrial Science and Technology (AIST) will submit a budget request for ¥30 million (\$200,000) to launch a 'microrobot' project next year, with the aim of developing tiny robots for the internal medical treatment and repair of human beings. The robots will also be used for the maintenance and repair of nuclear power plants and may find applications in space.

Although the initial budget request is small, as is usually the case with MITI projects in their first year, MITI is planning to pour ¥25,000 million (\$170 million) into the microrobot project over the next ten years under AIST's 'large-scale' (Ogata) industrial research programme.

The Ogata projects are among the biggest in AIST's repertoire, and there was stiff internal competition within MITI for selection of the one new project that can start next year. Furthermore, the proposed budget and project duration are unusually ambitious — Ogata projects typically run for seven years on budgets of about ¥15,000 million.

Kenzo Inagaki, deputy director of MITI's industrial machinery division,



Getting smart: a sketch of MITI's projected 'pill'. No delivery dates have yet been announced

which helped to formulate the project, says that although MITI's draft proposal discusses microrobots manipulating cells, robots that travel through blood vessels *à la* Spielberg are "at least a hundred years off". But he thinks that robots of about a centimetre in size that can enter the stomach and intestine are feasible.

One idea is to develop a 'smart pill' that will be equipped with a sampling device and sensors in the nose cone (to detect, for example, pH and temperature), and a

drug delivery system in the rear.

Several robot manufacturers, including FANUC and Fuji Electric Co., are expected to join the project, together with semiconductor manufacturers such as Hitachi, NEC and Toshiba. And Inagaki says that, although "officially" Ogata projects are totally financed by MITI, industry is expected to contribute "an additional 30 per cent or so".

The project was proposed by a MITI committee of industrialists and academics headed by Naomasa Nakajima of the department of mechanical engineering at Tokyo University. Another member of the committee is Isao Karube of the Research Center for Advanced Science and Technology of the same university, well known for his work on biosensors, such as blood glucose analysers.

Inagaki says that research in the field of micromachines is only about five years old and that the number of researchers worldwide with expertise is very limited. For this reason, MITI officials are hoping to make the project into a joint international effort. But they do not intend to

approach potential foreign collaborators until details of the project are finalized.

The project is an amalgamation of proposals and includes plans to develop microrobots for the maintenance and repair of nuclear power plants, chemical plants, gas and water pipes and aircraft engines. And MITI foresees a day when microrobots will be able to patrol the piping of nuclear power plants and carry out repairs "whenever an abnormality is detected", avoiding expensive shut-downs for maintenance and repair.

One likely addition to the project is the development of robots for use in space, already being investigated under another Ogata project.

MITI's budget proposal will no doubt be trimmed down in negotiations with the Ministry of Finance. And it remains to be seen if the Ministry of Health and Welfare will remain silent while MITI launches a project that delves deeply into the realms of medicine and drugs. But MITI will probably disguise and alter the project in such a way as to avoid objections.

David Swinbanks

CHEMICAL WEAPONS

Heated argument in prospect

Sydney & Boston

CRITICISM of the US Army's programme to incinerate chemical weapons on Johnston Island in the Pacific has come from several quarters in recent weeks, even though the incineration programme is already fully operational and more US chemical weapons are on their way to the island from bases in Germany. The debate, both in the United States and in the South Pacific countries, has caused a split between environmentalists and arms control advocates, but now seems unlikely to affect the fate of the Johnston Island programme. Nevertheless, it may well have an impact on the long-term prospects for incineration as a way for the superpowers to dispose of their vast and ageing stockpiles of chemical weapons.

At a meeting of the South Pacific Forum, held earlier this month in the small Pacific nation of Vanuatu, some participants continued to criticize the US chemical weapons incineration plant on Johnston Island as "another example of the Pacific being used by major weapons-producing states as an experimental area". Of the forum nations, only Australia and New Zealand supported the Johnston Atoll Chemical Agents Disposal System (JACADS), which began operation on 30 June (see *Nature* 346, 5; 5 July 1990). A communique released by the island nations opposing the plant complained about the "significant uncertainties and risks" posed by the project.

Nor were the forum participants reassured by three scientific reports presen-

ted by Australia's Prime Minister, Bob Hawke, in support of US claims that the weapons destruction methods include adequate environmental safeguards. According to the US Department of Defense, workers at the JACADS incineration facility have already safely destroyed more than 800 artillery shells and some 2 tons of nerve agent. By now, Johnston Island, 1300 km southwest of Hawaii, has accumulated some 300,000 weapons ready for incineration, including mustard and nerve gas shells. Another 100,000 US shells consisting of 400 tons of Sarin and VX nerve gas are in transit from Clausen, West Germany.

The environmental advocacy group Greenpeace continues to be one of the most vocal opponents of JACADS, claiming that the high-temperature incineration process produces toxic bioaccumulative chemicals such as polychlorinated dioxins and furans. According to Paul Johnston, Greenpeace International spokesman on chemical weapons, there are four safer alternatives: chemical neutralization; photodegradation; electrochemical procedures and biodegradation.

These methods, Greenpeace argues, have the advantage over incineration that they would be done in closed systems, and would not "result in huge volumes of gases being pumped into the atmosphere". A recent Greenpeace technical review* of the JACADS programme reviewed the

prospects and current status of each of these alternative methods. Johnston says that these alternatives could be developed into working techniques within 18 months if the US Army were to conduct accelerated research on them.

But Trevor Findlay, senior research fellow at the Peace Research Centre of the Australian National University in Canberra, does not believe that the alternatives offered by Greenpeace are totally safe either. As with incineration, Findlay points out, all four proposed alternatives require the lethal chemical agents to be taken out of their shells. The possibility of spills and accidents, he says, poses a safety and environmental threat from almost any proposed method.

Now that US Defense Secretary Dick Cheney and congressional committees have agreed in principle to cancel the US binary chemical weapons modernization programme, arms control advocates in the United States are heartened by US actions and see JACADS as "a big step forward" in the effort to live up to an agreement, reached with the Soviet Union last spring, that at least 80 per cent of both superpowers' chemical stockpiles should be destroyed. Many other countries, even some Pacific nations, have supported the demilitarization effort despite the potential environmental threat posed by chemical weapon destruction.

Australia, for instance, has supported JACADS as part of a push towards an international chemical weapons treaty (*Nature* 341, 271; 1989). Unless it destroys its own stockpile, the United States is seen to be in no position to push other countries into signing a treaty.

But, environmentalists point out, even with the addition of the US chemical weapons from Germany, only some 7 per cent of the US stockpile will be destroyed at the Johnston Island facility. The Soviet Union's stockpile, none of which has yet been destroyed, is believed to be larger still. Thus the fate of most of the superpowers' chemical arsenals remains uncertain, and the environmental stakes are high. The Soviet's plans are unclear, although they too are working on incineration methods and are committed to collaborating with the United States in the weapons' destruction.

The remainder of the US chemical weapons arsenal is marked for incineration at eight domestic facilities, yet to be built.

Environmentalists say that they will fight the plan and push for more research into alternative methods. "Our fight at Johnston Island is not over yet and our serious concerns continue about the environmental damage caused by incineration", says US Greenpeace spokesperson Sebia Hawkins; "the larger fight in the United States is just beginning".

Tania Ewing & Seth Shulman

* Greenpeace Review of Johnston Atoll Chemical Agent Disposal System (JACADS), July 9, 1990.

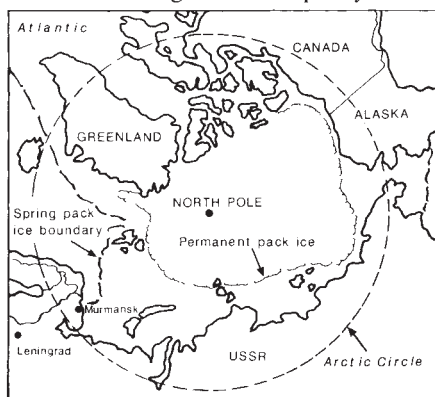
Soviets break the ice

Munich

PERESTROIKA and concern over global warming may combine to give Western researchers access to the Arctic through one of the most inhospitable ocean routes in the world. Taking a cue from a speech by Mikhail Gorbachev at Murmansk in 1987, the Soviet government earlier this summer commissioned an international feasibility study to determine the economic and ecological consequences of the year-round opening of the ice-choked passage (see map) known as the Northern Sea Route. Researchers from three Western institutes met at Oslo in late June to begin the study, which they expect to complete by March next year.

Controversial studies of global warming done by Peter Wadhams, a University of Cambridge researcher, and others, indicate that the sea ice in the Arctic regions may be thinning (see *Nature* 345, 762 & 795; 1990). Researchers are eager to obtain data concerning ice thickness, ice motion, and climate history, all of which might yield valuable clues about global warming, and it seems certain that access to the Soviet Arctic will increase.

"Soviet [polar] scientists have always been enthusiastic about cooperating with the West", says Wadhams. "Now it has become official government policy". Col-



laborations were in the past always severely limited by Soviet security concerns, he adds.

As part of a joint West German-Norwegian-European Communities project, Norway recently received permission to place in Soviet waters a sonar device that measures the thickness of sea ice, says Wadhams. More such devices might soon follow.

For the Soviet government, the Northern Sea Route is of considerable economic importance. During each summer, when icebreakers are used to clear a path, 1 to 4 million tonnes of cargo are shipped through the passage, which provides a short route from the Atlantic to the Pacific. The distance from Hamburg to Yokohama is reduced from about 12,000 nautical miles (through the Suez Canal) to

less than 7,000 nautical miles and the journey time from 30 to 22 days.

The Soviets have built a large fleet of nuclear-powered and conventional ice breakers, says Arnfinn Jørgensen-Dahl of the Fridtjof Nansen Institute in Oslo, and by using them to keep the passage open all year round they could help their own economy and earn hard currency by charging foreign ships a fee for the journey.

But there is concern among Western researchers about the possible danger to the environment of the Soviet project.

"We assume there are dangers, especially from an oil spill", says Jørgensen-Dahl. Even though it could not stop the Soviet Union from opening the passage, the West can make a "positive contribution", he says, by ensuring that environmental damage is limited.

According to Wadhams, however, the dangers are "not negligible but not extremely serious". He does not expect oil tankers to use the passage. The environmental impact of opening the passage to international traffic, he says, would not be "any more severe" than the impact of current Soviet shipping.

Steven Dickman

ECOLOGICAL SCIENCE

Return of life unobserved

Mount St Helens, Washington

TEN years after the 50-megaton volcanic eruption that scorched square miles of woodland into Mars-like desolation, Mount St Helens is coming back to life, largely unobserved. Most of the scientists who raced here in the summer of 1980 to record the first glimpses of life in this barren world are back at their universities, unable to get further funding for research at the volcano, and their minutely compiled data from the early years after the eruption now gather dust. But researchers say that, ten years on, Mount St Helens is perhaps more interesting than it was immediately after the explosion.

Its original ecosystem removed at a stroke, conditions at Mount St Helens approach laboratory purity. The species that first repopulated the blighted landscape in isolated pockets are now beginning to interact, giving a unique perspective on survival and evolutionary strategies. In the hot springs that flow from the crater itself are thermophilic bacteria that may resemble the first life on earth. And in the volcanic ash and mud, 200 metres deep in places, small patches of plant life constitute a textbook case of succession, one species paving the way for the next in a process that can take centuries elsewhere.

But as the attention of funding agencies has moved elsewhere, unique scientific opportunities at Mt St Helens are being lost. "It's a fleeting opportunity", says Robert Williams of the US Department of Agriculture. "Things aren't static, they won't be there for us next year". During the first four years after the explosion the recovery of Mt St Helens was documented in detail, but those baseline studies have been abandoned and the continued biological evolution, now proceeding at exponential pace, is being recorded in only anecdotal fashion.

Research funding, which between 1980 and 1984 was over \$1 million a year, is now less than \$100,000 annually. More than

100 scientists studied the region in the early years, but now fewer than a dozen biologists make occasional trips — often with time and money borrowed from other research projects. Although 25 scientists from the US Geological Survey still monitor the volcano for seismic activity, research on accelerated erosion in the valley created by the vast 1980 mudslides is now better supported by the local logging company than by federal agencies.

Frustrated by the loss of a scientific opportunity at Mt St Helens, a group of biologists, geologists and ecologists earlier this year proposed a new "Mt St Helens Initiative", a \$900,000-a-year scientific return to the region. The scientists hope to get money for their project next year.

But going from concept to an actual appropriation will take some salesmanship. For the Mt St Helens scientists, that may be the biggest challenge of all. Unlike physicists and astronomers, who have been forced to master the ways of Washington to gain funding for their expensive projects, the initiative's supporters, like many "small science" researchers, appear to be badly out of touch with the politics of science.

At a day-long helicopter tour of the volcano for congressional staff and journalists last week, the researchers fumbled their chance to make a direct appeal. Abstractions like "interdisciplinary studies across the disturbance gradient", "infrastructure" and "synthesis" may be full of appeal to environmental researchers, but the congressional staff were clearly lost. Acknowledging a tactical error, the researchers privately said that they would attempt to draft a more lucid proposal and send it to selected members of Congress within weeks. Given the modest request, some increased support seems likely. But how much more could be won with a clever pitch may remain a hypothetical question.

Christopher Anderson

BRITISH ASSOCIATION

A matter of education

Swansea

BRITAIN is "in danger of becoming one of the least adequately educated of all the advanced nations", according to Sir Claus Moser, president of the British Association (BA). In his address to the BA's 1990 meeting in Swansea earlier this week, Moser called for an urgent review of the whole of UK education and training, with the 1990s designated the 'decade for education'.

Moser welcomed the plan of Education Secretary John MacGregor to double UK participation in higher education over the next 25 years, noting that only 15 per cent of British school leavers now go on to higher education, compared with about half in the United States. But the expansion will need increased finance, Moser said, "of which so far there are no signs". Expansion of higher education must go hand-in-hand with a broadening of courses and the encouragement of under-represented groups such as women, ethnic minorities and the poorer section of the community, he added.

Moser's views are unlikely to be received favourably by a Conservative government determined to let market forces take the lead in educational policy.

Higher education reforms are a small

part of a wide-ranging review envisaged by Moser. The fact that most children leave school at the age of 16 is "the most unacceptable feature in [British] education", he said. Britain should adopt a system similar to the French *Baccalauréat* (a broad-based university entrance certificate), he urged. Moser also voiced concern over the quality of UK secondary school science teaching — a recent BA-sponsored survey found that one in five British science teachers believe they lack adequate knowledge to teach their subjects. Regular 'refresher courses' are needed to improve this situation, Moser said, with salary increases to boost morale and recruitment.

Moser blamed ingrained British attitudes for the state of the country's educational system, but singled out the present government for an erosion of the social sciences in the United Kingdom. He argued strongly that the social disciplines are true sciences, denouncing the government's removal of the word 'science' from the title of the Economic and Social Research Council in 1983. Moser emphasized the importance of the social sciences' interaction with the 'harder' sciences in tackling problems such as global warming.

Peter Aldhous

CARBON DIOXIDE SOURCE

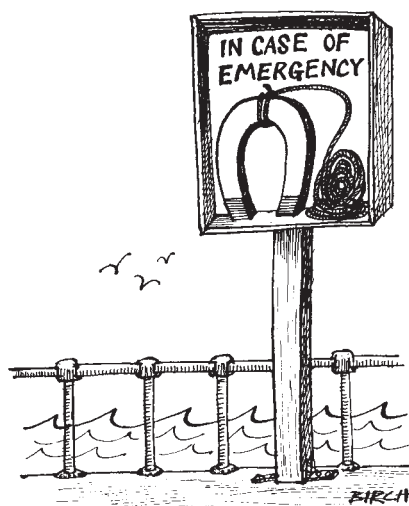
Filing plans for algae

London

PLANS to test the avant-garde idea, first mooted in the United States, that levels of atmospheric carbon dioxide could be reduced by sowing the oceans with tons of iron filings were unveiled by a marine biologist this week at the British Association for the Advancement of Science in Swansea.

Phillip Williamson of the Natural Environment Research Council's Plymouth Marine Laboratory announced plans for a "discreet" ocean test, which will be organized jointly by the Plymouth laboratory and the University of Rhode Island. The project would be smaller than the flagship proposal from a group of US researchers, which is currently under consideration by the US National Academy of Science, to deposit about a ton of iron over some 10,000 square kilometres of sea.

Williamson is one of a growing number of researchers who believe that in principle 'fertilizing' the sea with iron filings could boost the growth of carbon-dioxide absorbing marine algae sufficiently for significant amounts of the greenhouse gas to be sucked out of the atmosphere. Their belief stems from demonstrations of iron-induced algal growth in bottled sea water.



But the iron-filing idea has met with opposition from environmental groups such as Greenpeace who view it as a "bizarre technical fix" with unforeseen and potentially dangerous side effects. Its protagonists say that it is essential to find out now, whether iron filings could be used to bring down carbon dioxide levels in the event of a global warming emergency. And the only way to do that, they argue, is by conducting a small-scale test in the sea.

David Concar

BIOTECHNOLOGY

Cetus forced into staff cuts

San Francisco

IN what appears to portend a change in corporate strategy, Cetus, the California-based biotechnology company, last week announced the resignation of president and chief executive officer Robert A. Fildes and the lay-off of 100 of its 950 workers. The news comes just three weeks after the refusal by a US Food and Drug Administration (FDA) advisory panel to recommend for approval Cetus's prize cancer drug, interleukin-2 (IL-2) (see *Nature* 346, 501; 1990).

The company attributed Fildes's resignation on 16 August to "differing views regarding the company's priorities", which biotechnology analysts take to mean that company officials were unwilling to continue Fildes's strategy of risking much of the firm's future on the much-delayed FDA approval of IL-2. Ronald E. Cape, who has replaced Fildes as chief executive officer, indicated that Cetus will pursue "alliances" with other companies in order to build business and reduce its risks. A spokeswoman denied that Cetus is considering selling off any parts of the company.

Announcement of how the workforce cuts will be made is expected by mid-September. Meanwhile, company officials are still trying to schedule their next meeting with the FDA to determine what additional data on IL-2 will be required, or whether additional clinical trials are necessary, which would lead to an even longer delay. Analysts now estimate that approval of the drug will be delayed for 12 to 18 months.

Elizabeth Schaefer

CLEAN TECHNOLOGY

Councils to cooperate

Swansea

THE UK research councils have set up a unit to provide a focus for developing new products and processes that do not harm the environment.

The Clean Technology Unit, established by the Science and Engineering Research Council and the Agriculture and Food Research Council, will initially provide £1 million to top up 40 existing projects relevant to clean technology.

Announcing the formation of the unit at the British Association's conference in Swansea on Monday, its head Dr Nicholas Lawrence said this would rise to £10 million per year.

The first specific research projects will be announced in early 1991, and will focus on three areas: new processes that are inherently cleaner, modifications of existing processes to reduce or eliminate pollution and retrofits to deal with past pollution problems such as landfill sites.

Nuala Moran

A genuine global partnership?

S. S. Yamamoto

Japan is considering a huge financial investment in the US Superconducting Super Collider. But other types of contribution, to this and any international project, are equally important.

A HIGH-LEVEL delegation from the US Department of Energy recently visited Japan with a letter from President Bush to Prime Minister Kaifu soliciting about \$1.6 thousand million for the construction of the Superconducting Super Collider. The request was made in the spirit of the agreement between Bush and Kaifu to form a 'global partnership'. But I have serious doubts as to whether financial support of the Superconducting Super Collider falls into the category of global partnership. Although the United States invite foreign participation where money and manpower are concerned, to the best of my knowledge they do not invite participation on an equal basis in scientific, technical and administrative matters. By definition, a global partnership must involve such participation. If the past record of the US-Japan cooperative programme in high-energy physics is anything to go by, involvement of Japanese scientists will be secondary to the country's financial contribution in the case of the new project.

The US government is not to blame. The problem is Japan's. An impediment to genuine international collaboration is the reluctance on the part of the Japanese government to provide enough money for its physicists who work on collaborative research projects overseas, and the reluctance on the part of Japanese universities and research institutes to let their people, particularly senior people, spend an extended period overseas to participate fully in an experiment. Junior physicists and graduate students, some of whom are on the payroll of US institutions participating in the experiment, must bear the burden of carrying out the work without much guidance or leadership. This is detrimental not only to the education and training but also to the significance of the scientific contribution of Japanese teams.

Wrong priorities

I vividly remember the comments of Professor Robert Wilson, a former director of Fermilab, in his talk on international cooperation in high-energy physics at the 1986 meeting of the Japanese Physical Society in Tokyo. After lauding Japan for its active participation in collaborative experiments in the United States, but without stating whether Japanese physicists had made any significant scientific contributions, he said, "and I particularly welcome your money". This candid state-

ment succinctly reveals the state of many of the US-Japan collaborative experiments then in progress. The situation is not very different now.

Ninety-one Japanese physicists have agreed to participate in the Superconducting Super Collider project, but some of these are from small groups of one or two and not likely to participate in a significant way. I see no prospect for improvement. There are only a few postdoctoral fellowships provided by the Japan Society for the Promotion of Science to send young PhDs to work on international collaborative projects for a certain fixed period without tenured appointments. Therefore, it is likely that only a small fraction of the 91 physicists will be able to work on the project full-time. Under these circumstances, I feel the expenditure of nearly \$2 thousand million is actually detrimental to the sound development of high-energy physics as well as other scientific disciplines in Japan for the following reasons.

Japanese physicists will be paying an entrance fee to join the project without really being able to make significant scientific contribution. Under the guise of US-Japan collaboration, some Japanese high-energy physicists have deprived themselves of an opportunity to do their own experiments, and some graduate students have been deprived of opportunities to work closely with their supervisors. In addition, if a sizeable fraction of the rather few university-based high-energy physicists are engaged full-time in the Superconducting Super Collider, how will their graduate students obtain PhDs? How can the domestic high-energy physics programme continue adequately if \$1.6 thousand million is spent on the Superconducting Super Collider? To educate and train future physicists, the number of university-based high-energy physicists must be increased, and Japanese high-energy accelerators must remain operational and new machines built. Some Japanese physicists believe that in return for the contribution to the Superconducting Super Collider, all future high-energy accelerator projects in Japan, such as the Japan Linear Collider, should be abandoned, and high-energy physics experiments should be done abroad. Such an arrangement will have untold ill consequences for Japanese high-energy physics.

There is also the question of allocating funds for collaborative research projects

in other countries, notably for LEP at CERN. There is already a successful collaboration in a multinational project called OPAL at LEP involving physicists from the University of Tokyo. At present, seven of these are working at CERN for extended periods. Some have been there for many years, and the cost of supporting them is borne by the university from a specially allocated budget. The capital expenditure for this collaboration is limited to the detector cost, but the case of the Superconducting Super Collider may cause CERN to request contributions from Japan to its operating funds. To have more balanced global collaborations, I think some of the money for the Superconducting Super Collider should be spent on this and other future international collaborations.

Political aspects

Finally, because the United States asked for money before adequate discussion among Japanese physicists as to the desirability and extent of their participation, the Superconducting Super Collider case has begun to take on a political aspect. A project of this magnitude is bound to be political to some extent, and of course physicists alone cannot decide the extent of the Japanese contribution. But if political considerations overwhelm the scientific ones, the project will bode ill for future international collaboration.

So should Japan contribute to the Superconducting Super Collider? My answer is a qualified yes — qualified because if the Japanese contribution is to be mostly monetary, my answer is no. But as the project is deemed to have significant scientific merit worthy of an enormous investment by many distinguished physicists around the world, and because there is strong interest in it on the part of some of the Japanese high-energy physics community, my answer has to be yes. The US government and high-energy physics community should insist that the Japanese contribution is scientific as well as financial. Because large-scale international collaborations are inevitable in many fields of basic and applied research, it is imperative that this case should be carefully discussed by physicists and government officials of both countries. □

S. S. Yamamoto is in the Department of Physics, University of Tokyo, Bunkyo, Tokyo 113, Japan.

Problems with telescopes

SIR—I should like to comment on the problems with the Hubble Space Telescope (HST).

The primary mirrors of large modern telescopes are tested in autocollimation at their centres of focus, the error being determined by interferometric or Hartmann type analysis. Because primaries of short focal length are strongly aspheric (slightly hyperbolic for Ritchey-Chretien telescopes), the autocollimated image suffers from strong aberration, which is compensated during the test by a 'compensation' or 'null' system. This technique was invented by the English amateur Horace Dall in 1947¹.

The null system must be correct to very high precision, not only in its design and manufacture but also in its position in the test set-up, otherwise a systematic error arises which is spherical aberration in its classic form.

The testing of convex secondary mirrors, which cannot form real images, is a subject in its own right², but all test methods have tolerances which, if they are not respected, will lead to similar spherical aberration. But there is one test of secondaries (together with the

whole system) which effectively ensures that matching error cannot occur: this is the Pentaprism test. The Pentaprism test is essentially a one-dimensional substitute for the autocollimation test against a plane mirror. The Pentaprism deflects small parallel beams about 90° and has the property that the deflection is unaffected by inevitable rotation errors of the Pentaprism as it is moved across a diameter of the pupil. The Pentaprism test will ensure only that the two mirrors will in combination produce an image free of spherical aberration at the nominal Cassegrain focus. Various other tests exist for faults such as astigmatism and high spatial frequency errors ('ripple'). The Pentaprism test was probably first invented by Wetthauer and Brodun in 1920³, and has been systematically used by a few manufacturers, for example REOSC of Paris⁴ or Korhonen in the recently completed 2.5-m Nordic Optical Telescope. It was used in the United States in 1939 by Hendrix and Christie⁵ for Schmidt systems, and was described by Hochgraf in 1969⁶.

It seems surprising that this test, which is simple and cheap to set up (the telescope axis can be vertical or horizontal) was not applied to the HST; it would certainly have revealed the error.

But its use is by no means general in telescope-making, which explains why matching error, often very large, has been a common technical error. It affected the Canada-France-Hawaii Telescope, for which the spherical aberration was successfully corrected by bending the secondary.

The ESO New Technology Telescope (NTT) also had a matching error, which was provoked by a systematic error in the positioning of the compensation systems used to test the primary. Although this error was small (1.8 mm in a test distance of ~15.4 m) the wavefront error caused by the spherical aberration thus introduced was ~3,000 nm. This produced an image containing 100 per cent of the light in 0.71 arc seconds. Although this was outside the NTT's passive specification, we were able to correct it completely—to less than 0.1 arc sec, the general 'noise' level of the system—by the first (fixed) level of the active optics system⁷.

The Pentaprism test had been proposed by ESO to the manufacturer, Carl Zeiss, but was dropped by mutual agreement because a double test system (visible and infrared) gave a cross-check of the individual mirror tests. Had a matching error existed outside the dynamic range of the NTT's active optics system, it would have been revealed by this comparison. Further details will be given in the *ESO Messenger*, September 1990.

The HST, by contrast, has a very stiff, lightweight, egg-crate type of primary, and the dynamic range afforded by the 24 actuators is far too small to be able to correct an error in the primary of the magnitude found. Because the HST is effectively a passive telescope as far as correcting the primary is concerned, the Pentaprism test would have been the best guarantee against matching error. It is reliable, simple and cheap.

It has been said that the current performance of the HST is as good as the best ground-based telescopes. This is certainly not true. Apart from the NTT, which because of its active optics routinely produces images smaller than 0.5 arc second at the excellent site in La Silla, Chile, there are several passive telescopes, including the William Herschel Telescope in La Palma, capable of producing sub-arcsecond images under favourable seeing conditions. By contrast, the HST has an image size of at least 1 arcsecond from the spherical aberration alone.

RAYMOND N. WILSON

*European Southern Observatory,
Karl-Schwarzschildstrasse 2,
D-8046 Garching bei Muenchen, FRG*

1. Dall, H.E. *J. Br. Astron. Ass.* **57**, 5 (1947); *Amateur Telescope Making*, Vol. 3, 149 (Scientific American Inc., 1953).
2. Wilson, R.N. *ESO Tech. Rep. No. 3* (1974).
3. Wetthauer & Brodun *Z. Instrumentenkunde* **40**, 96 (1920).
4. Esparid, J. & Favre, B. *Nouvelle Rev. d'Optique Appliquee*, **1**(6), 395 (1970).
5. Hendrix, D.O. & Christie, W.H. *Scient. Am.* (August 1939).
6. Hochgraf *Opt. Sci. Center News*, January-March 1969, 41.
7. Wilson, R.N., Franza, F., Noethe, L. & Andreoni, G. *ESO Technical Preprint No. 24* (1990).

Foreign aid

SIR—The British Natural History Museum (NHM) can apparently no longer find the necessary £20 million or so annually to fund its present level of activities, much less pay for needed renewals and expansions. Yet it is an institution "of great international importance", its collections belonging "to the world, not just to London or the British Isles" (Adam Urbanek and Michael Novack; *Nature* **345**, 656; 1990). This suggests a solution to the present crisis. If the United Kingdom is now too poor to finance the NHM adequately, let us rename it the International Museum of Natural History, and arrange to support it with funds from the International Union of Biological Sciences, the Smithsonian Institution and any other international or national organizations and individuals who recognize the necessity for such action at the present time.

Meanwhile, Britain feels that it can afford Trident missiles, a sad indication of where its priorities lie today.

RALPH A. LEWIN

*University of California, San Diego,
Scripps Institution of Oceanography,
La Jolla, California 92093, USA*

Correction

In the opening paragraph of his letter about the Natural History Museum, Des Griffin (*Nature* **346**, 99; 1990) was referring to commentary on the subject in general rather than to Sir Walter Bodmer's Commentary in *Nature*. □

Need for religion

SIR—The effort expended by scientists to prove God to be a myth and religion to be a superstition would be more fruitfully spent in finding out why belief in transcendental power is so widespread: just about every culture includes a religion. I can think of two reasons:

- (1) Religion helps societies to function more effectively, giving them a competitive edge over atheists.
- (2) Some humans are dimly aware of another dimension of this Universe, or even of some other continuum, that they call God. This awareness helps them make more constructive decisions than those of people lacking such a faculty.

A simplistic view indicates that, in the struggle for survival, atheists should have prevailed, as religious practices are a drain on resources. Yet it is the atheistic societies that failed to survive. Religion must have something going for it. Instead of pretending that it does not exist, we should try to see what makes it tick.

ANDREJS Baidins

*1104 Windon Drive,
Wilmington,
Delaware 19803, USA*

Assertive sentence titles

SIR—If it is such a bad thing for molecular biologists to use declarative sentences in journal titles as Rosner asserts¹, why have they made so much progress, at a time (1982–89) when this deplorable process so increased in popularity that only *Nature* seems to have escaped from the floodtide of assertive sentence titles (ASTs)?

Is it the result of a rise in confidence in a process which is so successful that buyouts and allegations of insider dealings in biotech companies are as much news in *Nature* as scientific budgets²?

Or has success so emboldened biologists to test their work by the methodology of proof by falsification that they have confidence to use declarative English? And are not declarative sentences much easier to prove erroneous than open-ended questions, which really are akin to playing tennis with the net down — no rules need apply? An example will suffice.

Rosner erroneously asserts that ASTs are often patently unprovable, for example declarations which cite that a process requires a certain component; Rosner claims that only when all possible conditions are tried can they be substantiated. But in fact, only the negation of the condition need be demonstrated. Declarations that a process requires or does not require a certain component are substantiated through the conspicuous use of control experiments.

The withdrawal by Fleischmann and Pons of their cold fusion paper submitted to *Nature*³ had less to do with the absence of title citations in this journal's references, as one could infer from Rosner's hypothesis, than the referees' criticism of the need to control for possible neutron contamination from other sources, such as cosmic rays⁴. The cold fusion account from the Brigham Young group was published⁵ after it corrected for possible background contamination in the results.

ASTs make for progress only when the underlying papers survive the same process of scrutiny used to test scientific hypotheses in other disciplines. It is the open-ended nature of the falsifiability criterion and not grammatical constructs that make for the tentativeness of scientific assertion and keep the scientific endeavour a process and not a product.

EDWARD J. KROWITZ

2415 North Dickerson Street,
Arlington,
Virginia 22207, USA

1. Rosner, J.L. *Nature* **345**, 108 (1990).
2. Gershon, D., *Nature* **345**, 102 (1990).
3. *Nature* **338**, 691 (1989).
4. Carpenter, J.M. *Nature* **338**, 711 (1989).
5. Jones, S.E. et al. *Nature* **338**, 737–740 (1989).

Relativity values

SIR—Stefan Marinov's apparatus designed to test relativity (*Nature* **346**, 103; 1990) is clearly similar in principle to the electromagnetic flow-meter, which works by measuring the voltage induced by fluid moving in a magnetic field.

Assuming that such an instrument is set up to give zero output for zero flow, a reversal of flow direction reverses the polarity of the output. I would therefore say that "the polarity of the slider depends on the direction of the current in the concentric circuit, and also on whether the slider moves clockwise or anticlockwise".

We use electromagnetic flow-meters both as current-meters, in which case the instrument is fixed and the water moves, and as ships' logs, where the water may be stationary and the instrument moves. However, I have not yet managed to get an output from such an instrument by putting it in a bucket of water and marching across the laboratory with it.

A. J. BUNTING

Institute of Oceanographic Sciences,
Deacon Laboratory,
Brook Road, Wormley,
Godalming,
Surrey GU8 5UB, UK

Metric false trail

SIR—The influence of the metric system both in science and in everyday life has also been responsible for a fatal flaw in conventional computation. This is that the binary code of digital computers is exponentiated in radix ten numbers, instead of base two numbers. The result is that every switching operation carried out at the system clock frequency of a computer is essentially "bringing down the exponent" in order to carry out conversions between the respective number bases. All information is expressible in bits (binary digits) and there is therefore no reason why computation cannot be carried out directly in the binary system. The ineluctable advance of the metric system supposed by your correspondent C. H. Evans (*Nature* **345**, 658; 1990) will come to a permanent halt as soon as it is realized that the Emperor at the head of the procession has no clothes.

B. E. P. CLEMENT

Everest Drive,
Crickhowell,
Powys, NP8 1DH, UK

Never again

SIR—I should like to comment on the review by Robert N. Proctor (*Nature* **344**, 502; 1990) of Paul Weindling's book *Race and German Politics between National Unification and Nazism 1870–1945*.

J. F. Lehmanns Medizinische Buch-

handlung GmbH was founded in 1981 as an independent company; in the same year, the five Rothacker bookshops were bought by the Deutsche Ärzte Verlag. The name J. F. Lehmanns was chosen for these bookshops without anyone being aware of the infamous past of this name.

For many years after the Second World War, high-level and specialist medical books with a high reputation in scientific circles were published under the imprint J. F. Lehmanns. This publisher has since been amalgamated into the well-known company Springer Verlag.

It was certainly a mistake not to research the background of this name. The past nine years have, however, proved that our company and its employees cannot be associated with any thought of nazism, however expressed. As far as it is in our power to determine, fascism and other extreme ideologies will never again stand a chance with us.

MANFRED HAHN

J. F. Lehmanns,
Universitätsstrasse 31,
D-8400 Regensburg 1, FRG

BSE gene-linked?

SIR—Managed cattle breeding, progeny testing and sire selection are commonplaces of British livestock husbandry. As a result, many of our herds have considerable common ancestry; and with widespread and cheap artificial insemination, it is not impossible for one bull to impregnate tens of thousands of cows, many of them in successive generations.

Bovine spongiform encephalopathy (BSE) is a peculiar disorder. It appeared abruptly, nationally and with no obvious geographical foci. If it is caused by an infectious agent, then this is both difficult to detect and difficult reliably to transmit. Indeed, those animals to which it is irregularly transmissible appear to come from the epidemiologically 'at risk' group. We see it as a transmissible disease primarily by analogy with scrapie and other neuropathologies; and we have sought an explanation for the rapid onset of diagnosis in the use of offal in stock feeds. One could, however, also see BSE as a genetic disorder brought to prominence by the extensive inbreeding and commonality of germ lines.

In view of this, it would be interesting if a researcher with access to the appropriate databases were to run a check on the ancestry of those cows diagnosed as having BSE. If BSE were to be shown to be the outcome of inbreeding rather than the transmission of an infectious agent, the economic implications — and the policy measures that should be followed — would have implications on a national scale.

O. H. G. SPARROW

35 Bradbourne Street,
London SW6 3TF, UK

Towards a history of the Sun?

There seems to be just a chance that geochemists could reconstruct a record of the Sun's activity as a main-sequence star from measurements of the abundance of xenon-126 in rocks.

HERE is a conundrum for climate modelers and for the global warming community in particular: if it is true, as astrophysicists say, that the luminosity of the Sun has increased by 40 per cent in the past 5,000 million years, how is it possible to explain that the Earth has only just (since 18,000 years ago) emerged from a major glaciation? The question is not new; it has been on people's minds for decades. The need to answer it has been made compelling because the climate models that predict global warming have not yet been turned systematically and successfully onto historical data, with the consequence that the sceptics of global warming can fairly jeer: "If you cannot explain why there was, say, a Little Ice Age in Europe in the eighteenth century, why should we take seriously your predictions of the future?"

The postdiction problem is not, of course, as simple as it sounds. First, there is the question of what data to use. Ice-core measurements may have helped to define the composition of the atmosphere during relatively recent climate fluctuations, but the data are necessarily less precise than from measurements of the modern atmosphere. Estimating the composition of the Archaean atmosphere is, by extension, hardly distinguishable from guesswork.

The input of energy from the Sun is another imponderable. Variations of the Earth's orbital elements can be calculated well enough, at least over intervals of a few million years, to be included in climate models. But while satellite measurements of the solar constant suggest that fluctuations of the output of energy from the Sun are too small to account even for the barely distinguishable climatic trends in recent decades, it is always possible that earlier (larger) climatic fluctuations were somehow linked with variation of solar output — the coincidence of low winter temperatures in eighteenth century Europe and the Maunder sunspot minimum often seems too good to be true.

Over much longer intervals, there seems no doubt that the solar constant has changed, and indeed increased substantially. The Sun has been a main-sequence star for at least 4,500 million years, during which time hydrogen has been converted into helium in the core of the Sun, thus increasing the opacity of the material of the core, the central temperature and, eventually, the luminosity.

Astrophysicists will say that the structure and, indeed, the evolution of the Sun is one of the fixed certainties in their otherwise uncertain world. That, of course, is why there has been such consternation in the past 20 years because the number of neutrinos reaching the surface of the Earth is smaller (by between a half and two-thirds, according to energy) than the calculations suggest. Another difficulty is that, if there has been a 40 per cent increase of the solar constant in 5,000 million years, why have we only just emerged from an ice age — a period much colder than at the end of the Cretaceous, for example?

Despite astrophysicists' confidence in their picture of the evolution of the Sun, there are of course ways in which they might have been misled. One of the underlying assumptions in the astrophysical model of the Sun, for example, is that the hydrogen-burning core is thermally stable, or that convection is not an important cause of mixing. If the assumption is incorrect, there could well have been uncounted variations of solar constant in the past 5,000 million years.

Whence the ambition of W. C. Haxton, from the University of Washington at Seattle, to find a way of estimating the time-course of solar luminosity from measurements on the surface of the Earth (*Phys. Rev. Lett.* **65**, 809; 1990). How could that be possible? Only if the terrestrial consequences of the solar neutrino flux could be significant. That drives Haxton to geochemistry. His ingenious proposal is for a scheme for the measurement of the past neutrino flux from the Sun by means of the conversion of tellurium into xenon isotopes under the influence of energetic solar neutrinos.

That solar neutrinos should have had terrestrial consequences should be no surprise. The known deficiency of solar neutrinos was, after all, first recognized by the conversion of chlorine into neon isotopes by solar neutrinos. Since then, much ingenuity has been spent on other schemes for using terrestrial isotope abundances as measures of the neutrino flux, both as detectors for measuring the contemporary flux and as geochemical proxies for the past abundance of neutrinos. Haxton points out that there is already a geochemical scheme for the measurement of past neutrino fluxes based on the isotopes of technetium (element number 43, which has no stable isotopes). Indeed, ^{98}Tc is produced exclus-

ively by neutrino interactions, so that its concentration at any time could be a measure of neutrino flux over periods of time comparable with the half-life of the isotope. The obvious snag is that the technetium isotopes are so rare as to be hardly measurable and are, in any case, unstable.

Haxton's scheme is based on a search through the table of isotopes for a material likely to be made almost exclusively by neutrino interaction and which, although itself unstable, will yield a stable daughter product. His suggestion is that the conversion by neutrino interaction of tellurium-126 (^{126}Te) into xenon-126 (^{126}Xe) may be just enough to provide a yardstick of past solar activity. Briefly, sufficiently energetic neutrinos will convert ^{126}Te nuclei into ^{126}I nuclei, a proportion of which then decay into stable ^{126}Xe , the rarest of the xenon isotopes (with a natural abundance of 0.089 per cent).

The outcome is a proposal for searching for excess abundance of ^{126}Xe in terrestrial rocks containing tellurium, gold ores for example, and using that to infer the past flux of solar neutrinos with energy above the reaction threshold (which would exclude all solar sources of neutrinos except the most energetic, those derived from the decay of ^8B). In principle, it should be possible to use the ^{126}Xe concentration of tellurides as a measure of the total solar neutrino flux since the formation of the rocks concerned, or the time at which they were last molten. The comparison of results from rocks with different ages could then, in principle, yield an estimate of the time-course of the neutrino flux.

Haxton is the first to acknowledge that the task will not be easy. An assay based on the measured isotope excess, estimated at 0.1 per cent, of the rarest isotope of a rare gas is hardly likely to be everybody's idea of a simple project. Indeed, although Haxton points out that ^{126}Xe is shielded from contamination by cosmic-ray neutron conversion of nearby isotopes by their rarity (and instability), the background problem is bound to be horrendous. But, luckily, geochemistry is sustained by the zealous enthusiasm of a handful of people, typified by Edward Anders at Chicago, who regard problems such as these as interesting challenges. It will be great fun if Haxton's ingenious scheme turns out to be a usable guide to the history of the Sun's evolution.

John Maddox

GAPs in understanding Ras

Michael H. Wigler

RAS proteins play a central role in neoplasia and growth control, yet we do not understand how they are controlled or what they control. The papers by Downward *et al.*¹ and Zhang and colleagues² on pages 719 and 754 of this issue indicate that the enigma of Ras may be yielding.

Mammalian *ras* genes were first discovered as the oncogenes of acutely transforming retroviruses³. Subsequently, *ras* genes with activating mutations were found in many human and rodent tumours, providing the first solid evidence that a common biochemical defect might underlie cell proliferation, or neoplasia⁴. The importance of *ras* in the scheme of growth control and neoplasia is further underscored by observations that *ras* function is required for many other oncoproteins, such as tyrosine kinases, and growth stimulants, such as the tumour-promoting phorbol esters, to have effect⁵. Moreover, *ras* genes are evolutionarily the most highly conserved oncogenes yet found, and occur in all eukaryotic organisms where they have been sought. There are great similarities between the function and regulation of *ras* in microorganisms and vertebrates, but this oncogene's main function in the yeast *Saccharomyces cerevisiae*, which is the stimulation of adenyl cyclase, does not apparently parallel its function in vertebrates^{6,7}.

The *ras* genes encode a family of small, closely related proteins found at the inner surface of the plasma membrane where they can bind guanine nucleotides and slowly hydrolyse GTP to GDP⁸. Only when they are bound to GTP can the proteins fulfil their role in cell proliferation^{8,9}. Rates of GTP hydrolysis by wild-type Ras protein, but not oncogenic protein, are vastly elevated *in vivo* relative to rates measured for purified Ras proteins *in vitro*⁹. These discrepancies led to the discovery of GAP, a GTPase-activating protein found ubiquitously in cells. GAP is a large protein with relative molecular mass 110,000 (110K). The carboxy-terminal 40K acts catalytically upon wild-type Ras proteins to stimulate their hydrolysis of bound GTP, but fails to stimulate GTP hydrolysis by mutant oncogenic Ras¹⁰.

Regulation

GAP is thus an excellent candidate to be an upstream regulator of Ras protein, capable of downregulating Ras, thereby keeping it inactivated in the GDP-bound state. Oncogenic Ras proteins would escape this regulation. The reports in this issue^{1,2} strongly support this notion. First, Zhang *et al.*² report that transfection of the GAP gene inhibits morphological transformation in NIH 3T3 indicator cells by

the normal H-*ras* gene, but not by the activated mutant v-*ras* gene. These results provide the first direct evidence that GAP can inhibit Ras function in mammalian cells. It was already known that GAP has sequence similarities to the *IRA* gene products of *S. cerevisiae*, which down-regulate yeast *ras* genes¹¹. Mammalian GAP, if expressed in yeast, can complement the loss of the *IRA-1* gene, and can inhibit the wild-type mammalian H-*ras* but not the activated oncogenic form when coexpressed in yeast¹².

Downward *et al.*¹ provide a deeper insight into the factors that regulate Ras function. They report for the first time a physiologically relevant event that leads to activation of Ras: stimulation of the immune system's T-cell receptor. The amount of Ras in the GTP-bound state in T cells rapidly rises tenfold when cells are stimulated with PHA or a monoclonal antibody directed to the T-cell receptor. Stimulation of the T-cell receptor is thought to activate phospholipase C, which can generate at least two intracellular messengers: diacylglycerol, which activates protein kinase C, and inositol phosphates, which can mobilize intracellular calcium. Downward *et al.* demonstrate that the activation of protein kinase C by the tumour-promoting phorbol esters increases even more rapidly the proportion of Ras bound to GTP in T cells.

Similar but lesser effects were observed in other cell types. In principle, an increase in the ratio of Ras bound with GTP to Ras bound with GDP could result either from an increase in the rate of nucleotide exchange on Ras or from an inhibition of GTP hydrolysis, owing, for example, to an inhibition of GAP activity. In fact, Downward *et al.* observed no alteration in the rate of nucleotide exchange on Ras upon treatment with phorbol ester, but they found a rapid decrease in GAP activity in treated cells, suggesting that a GAP-like protein, lying in the signal transduction pathways of T-cell stimulation and protein kinase C activation, regulates Ras.

The question of the placement of GAP in the Ras signal transduction pathway is not so easily resolved. Mutant forms of Ras which fail to have effector function also fail to interact properly with GAP, leading to the hypothesis that GAP may indeed be a target of Ras action^{13,14}. This remained an intriguing but weak hypothesis until recently, when Yatani *et al.*¹⁵ reported that GAP and Ras seem to cooperate in the inhibition of coupling between muscarinic receptors and atrial potassium channels. Each protein, purified, can

inhibit this coupling alone, as can be demonstrated by an *in vitro* patch-clamp preparation, and antibodies to either component can block the activity of the other. Yatani *et al.* speculate that a Ras·GAP complex may alter membrane phospholipid metabolism.

Effector function

Do the results of Yatani *et al.* conflict with those of Zhang *et al.*? Not necessarily. The latter are compatible with two possibilities. First, endogenous levels of GAP in cells may not be limiting for Ras effector function. High levels of exogenous GAP might decrease Ras function if the amount of GAP-bound Ras exceeds the levels of a stoichiometric target for GAP. Second, Ras might have more than one effector function in cells. It is clear from the comparative studies that Ras has had at least two effector functions in the course of evolution. Moreover, evidence has accumulated that Ras has more than one primary effector function in *S. cerevisiae*¹⁶. These observations serve as a warning that Ras effector functions in mammalian cells are likely to be complex.

Hope for the resolution of the function of GAP has flared anew with the riveting report by Xu *et al.*¹⁷ that *NF-1*, the locus involved in hereditary neurofibromatosis, encodes a protein that is similar to GAP and to the yeast *IRA-1* gene product¹⁸ (see also Ponder's News and Views article on page 703 of this issue). The similarity between the *NF-1* gene product and GAP is restricted to the domain that catalyses accelerated GTP hydrolysis by Ras. But, the similarity between *NF-1* and *IRA-1* extends well beyond this region, and *NF-1* is consequently closer to *IRA-1* than it is to mammalian GAP. This result suggests that there may be a family of GAP-like molecules which may interact with Ras.

Patients with a single copy of the *NF-1* disease locus are prone to sporadic benign growths of neuroectodermal origin, and hence, if hereditary neurofibromatosis

- Downward, J., Graves, J.D., Warne, P.H., Rayter, S. & Cantrell, D.A. *Nature* **346**, 719–723 (1990).
- Zhang, K. *et al.* *Nature* **346**, 754–756 (1990).
- Ellis, R. *et al.* *Nature* **292**, 506–511 (1981).
- Barbacid, M. *A. Rev. Biochem.* **56**, 779–827 (1987).
- Yu, C., Tsai, M. & Stacey, D. *Cell* **52**, 63–71 (1988).
- Birchmeier, C. *et al.* *Cell* **43**, 615–621 (1985).
- Beckner, S.K., Hattori, S. & Shih, T.Y. *Nature* **317**, 71–73 (1985).
- Field, J., Broek, D., Kataoka, T. & Wigler, M. *Molec. cell. Biol.* **7**, 2128–2133 (1987).
- Trahey, M. & McCormick, F. *Science* **238**, 542–545 (1987).
- Marshall, M.S. *et al.* *EMBO J.* **8**, 1105–1110 (1989).
- Tanaka, K., Matsumoto, K. and Toh-e, A. *Molec. cell. Biol.* **9**, 757–768 (1989).
- Ballester, R. *et al.* *Cell* **59**, 681–686 (1989).
- Adari, H., Lowy, D.R., Wilmsen, B.M., Der, C.J. & McCormick, F. *Science* **240**, 418–521 (1988).
- Cales, C., Hancock, J.R., Marshall, C.J. & Hall, A. *Nature* **332**, 548–551 (1988).
- Yatani, A. *et al.* *Cell* **61**, 769–776 (1990).
- Wigler, M. *et al.* *Cold Spring Harb. Sym. quant. Biol.* **53**, 649–655 (1988).
- Xu, G. *et al.* *Cell* **62**, 599–608 (1990).
- Buchberg, A.M., Cleveland, L.S., Jenkins, N.A. & Copeland, N.G. *Nature* (in the press).
- Bar-Sagi, D. & Feramisco, J. *Cell* **42**, 841–848 (1985).
- Noda, M. *et al.* *Nature* **318**, 73–75 (1985).

follows the pattern of hereditary retinoblastoma, the *NF-1* gene may be an anti-oncogene. If its product holds in check the activity of Ras, or a Ras-like protein, the sporadic loss of the good allele in the cells of *NF-1* patients would then lead to aberrations in the growth control of these cells due to the constitutive activation of Ras. In this interpretation, the *NF-1* gene product acts as an upstream negative regulator of Ras.

There is, however, a contradictory

interpretation. In some cells of neuroectodermal origin, Ras stimulates differentiation and blocks proliferation^{19,20}. If *NF-1* encodes the target for Ras action in neuroectodermal cells, the loss of this product might also lead to uncontrolled cellular proliferation. Thus, for now, the striking news about *NF-1* is perfectly ambiguous. □

Michael H. Wigler is at Cold Spring Harbor Laboratory, Cold Spring Harbor, New York 11724, USA.

SILICON CLUSTERS

All surface and no activity

Tony Stace

ELEMENTARY considerations of solid geometry show that as the radius of a sphere is reduced, the ratio of surface area to volume increases. If this picture is extrapolated to an atomic scale it is easy to appreciate that for a small collection of atoms — a cluster — such as Si_{21} , the majority of atoms reside on the surface. Given such circumstances it is, therefore, surprising to note that in a recent study of silicon clusters, Jarrold *et al.*¹ observed the clusters to be far less reactive for molecular oxygen than the surface of bulk silicon.

The technical importance of silicon in the fabrication of electronic devices has meant that processes like the oxidation of well characterized surfaces have been thoroughly investigated. With the trend towards microelectronics and the possibility of constructing nanometre-scale (10^{-9} m) devices², there is clearly a need to identify how the chemical and physical properties of a bulk metal or semiconductor change when reduced to atomic proportions; the study of clusters could fulfil that requirement.

Clusters from refractory materials, such as silicon, are best prepared using laser vaporization; an intense pulse of laser radiation strikes a metal surface, vaporizing material which is then caught in the flow of a buffer gas such as helium. The gas cools the vapour, and promotes nucleation, resulting in a selection of clusters of various sizes. As will be seen, the efficiency of the cooling process could have a key role in determining the reactivity of the clusters.

A complicating factor in studying reactivity is that the comparisons are made between ionic clusters and neutral surfaces of bulk silicon. Experimental convenience dictates that ionic clusters are used; the electric charge means that

clusters can be mass-selected and guided to the reaction zone. In the event, delocalization of the charge diminishes the magnitude of any electrostatic interactions between a cluster and a reactant molecule, and calculations suggest that there is very little difference between the structures of

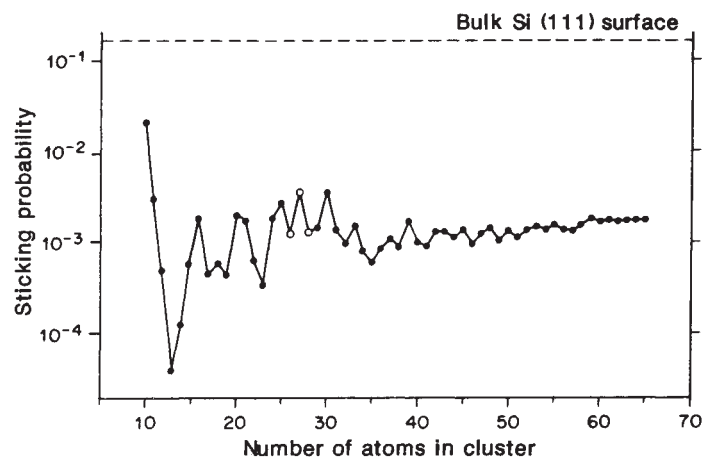
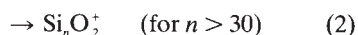
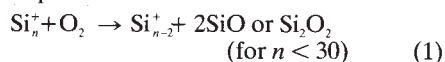


FIG. 1 Estimated sticking probability for the collision between an Si_n^+ cluster and an oxygen molecule as a function of cluster size. Shown as a dashed line is the result for the bulk Si(111) surface. (From ref. 1.)

large neutral and ionic silicon clusters.

Similar experimental considerations apply to the reaction products: charged products are readily identified using mass spectrometry, whereas the nature of neutral fragments can often be inferred from the thermochemistry. Thus, in the case of a reaction between a silicon cluster ion and an oxygen molecule, the dominant steps are¹:



The absence of detailed information on any neutral reaction products, together with the observation of step (2), supports the approach taken by Jarrold *et al.*¹ in interpreting the experimental data in terms of sticking probabilities. On this basis, a direct comparison with bulk silicon³ suggests that the probability of O_2

sticking on a cluster is lower by a factor of 100. Although the reactivity of smaller clusters with oxygen varies significantly with size (Fig. 1), beyond Si_{25}^+ the sticking probability is almost constant as a function of size.

Jarrold⁴ recently compiled a list of other molecules, such as methanol and ammonia, whose reactivity varies markedly with the size of the target silicon clusters. Possibly the most dramatic examples have come from the work of Smalley and co-workers⁵, who found that the reactivities of the clusters Si_{30}^+ and Si_{45}^+ with ammonia and ethene are several orders of magnitude lower than those of clusters one atom smaller or larger.

In a novel extension to those experiments, Smalley's group has shown⁶ that the reactivity of a silicon cluster ion can be modified through laser-induced annealing. In particular, annealing renders Si_{30}^+ and Si_{45}^+ virtually inert with respect to ethene and ammonia. But note that Jarrold *et al.*¹ did not see a similar discrimination on the part of these large clusters against oxygen. Such marked differences in behaviour between what are in essence identical experiments suggests that it may be necessary to consider in detail the way the silicon clusters are produced.

Laser vaporization leads to the generation of a plasma consisting of atoms, ions and electrons. The gas pulse which entrains the vaporized plume quenches the plasma and initiates nucleation. But because the binding energy between silicon atoms is quite high (the heat of atomization is about 4.6 electron volts), it is possible to produce 'stable' clusters with quite high internal temperatures, and only through further collisions and/or evaporation will they become annealed.

Depending on the exact conditions, three types of silicon cluster could emerge from the vaporization process: hot and possibly liquid-like; annealed, but to a metastable structure; or annealed to the ground-state structure.

Clusters from the first option might behave in a manner analogous to a hot surface, having low sticking probabilities owing to the rapid desorption of molecules. Beck and Andrews⁷ recently provided evidence of a laser-induced phase transition between solid-like and liquid-like forms of silicon clusters. Bearing in mind that the melting point of a cluster could be at least a factor of two lower than that of the bulk material⁸, estimates of the (photon) energy required to 'melt' a cluster may need to be re-evaluated. The possibility of the annealed options suggests the existence of structural isomers. These

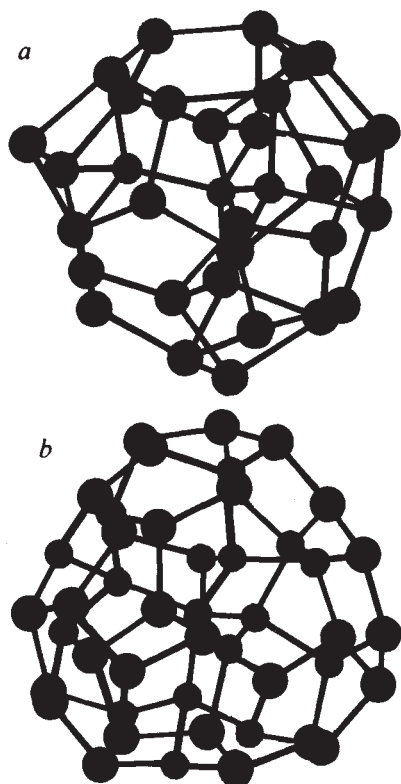


FIG. 2 Perspective views of Si_{33} (a) and Si_{45} (b). (From ref. 10.)

by several groups (see, for example, ref. 9) with reference to changes in the reactivity of niobium clusters.

It is, therefore, apparent that experimental conditions could strongly influence the type of reactive behaviour observed with metal and semiconductor clusters. Given that sufficient annealing does produce 'crystals' of silicon, it is clear from the results of Smalley and co-workers^{5,6} that certain atomic combinations are very much less reactive than their immediate neighbours.

To account for such behaviour, Kaxiras¹⁰ has suggested that certain combinations of silicon atoms, such as Si_{33} and Si_{45} adopt stable 'magic-number' structures (Fig. 2) which closely match the reconstructions observed for the silicon (111) surface in the bulk phase. The driving force behind surface reconstruction is the elimination of dangling bonds (unsaturated valencies), and it is a feature particularly

characteristic of semiconductor surfaces.

But having identified stable structures, we appear to be left with a conundrum: if Si_{33} or Si_{45} have very high surface areas and the morphologies of those areas are thought to resemble the bulk silicon (111) surface, why is it that the clusters are orders of magnitude less reactive than the

bulk? Could it be that clusters emerge from the vaporization-annealing process with internal temperatures that are higher than have previously been considered appropriate? □

Tony Stace is in the School of Chemistry and Molecular Sciences, University of Sussex, Brighton BN1 9QJ, UK.

MEMORY

Maintenance is presynaptic

T. V. P. Bliss

LONG-term potentiation, a sustained enhancement of synaptic efficacy induced by high-frequency electrical stimulation of certain pathways in the neural mammalian brain, may well provide the neural basis for at least some aspects of memory. Two reports, one by Malinow and Tsien¹ and the other by Bekkers and Stevens on page 724 of this issue², present compelling new evidence that long-term potentiation is expressed as a persistent increase in neurotransmitter release. This result may clarify the long-standing debate about which side of the synapse is responsible for the maintenance of long-term potentiation, and will direct attention to the cellular machinery by which long-term changes in neurotransmitter release can be achieved.

The trigger for long-term potentiation (LTP) is almost certainly the postsynaptic entry of calcium through the voltage-dependent channels controlled by the *N*-methyl-D-aspartate (NMDA) subtype of glutamate receptor³, a mechanism that provides a satisfying molecular explanation for just those properties of LTP that enhance its appeal as a memory model, namely associativity and specificity. By contrast, there has never been agreement on whether the locus of LTP expression is pre- or postsynaptic. The fact that induction reflects postsynaptic events does not by itself entail a postsynaptic location for the mechanism ultimately responsible for the sustained increase in synaptic efficacy. The debate has continued for many years between those favouring a postsynaptic mechanism, involving say changes in receptor number or sensitivity⁴, and those, myself included, who prefer a presynaptic mechanism expressed as a sustained increase in transmitter release⁵, with a position on the fence being a tenable and indeed often prudent option.

The latter two positions were made considerably less comfortable by reports from two groups in 1988 that the two components of the synaptic response, mediated by glutamate receptors of the kainate/quisqualate and NMDA subtypes, are differentially affected in LTP, there being little if any potentiation of the NMDA component^{6,7}. This observation seemed to favour a postsynaptic mechanism,

because an increase in transmitter release would be expected to affect both components, and indeed does so in paired-pulse facilitation and in post-tetanic potentiation (a short-term effect which, unlike LTP, occurs in most synapses). Proponents of a presynaptic mechanism were forced to search for *ad hoc* explanations such as a persistent downregulation of NMDA receptor/channel function following the induction of LTP. Now Malinow and Tsien¹ and Bekkers and Stevens² swing the pendulum towards the presynaptic side. These authors have applied the technique of whole-cell recording with patch electrodes to a quantal analysis of LTP in the hippocampal slice, and conclude that LTP is mainly presynaptic in origin.

Quantal analysis has long been the preferred way of establishing whether changes in synaptic efficacy are presynaptic or postsynaptic. The method exploits the discrete and probabilistic nature of the exocytosis of transmitter from vesicular stores at synaptic release sites to derive estimates of the binomial parameters *N* (number of release sites), *P* (probability of a nerve impulse leading to release from a given release site) and *a* (the synaptic response produced by a single quantum, usually identified with the contents of a single vesicle). *N* and *P* are presynaptic parameters, whereas *a* is postsynaptic (assuming constant packing of transmitter in vesicles). Analysis of changes in *N*, *P* and *a* following the induction of LTP can in principle indicate the locus of change.

The problem is that, unlike the neuromuscular junction with its single synapse, neurons receive thousands of connections at varying distances from the recording site in the cell body. This, together with the poor signal-to-noise ratio obtained with the high-resistance electrodes required to penetrate small cortical cells, greatly complicates the analysis⁸, and prevented general acceptance of an earlier quantal analysis⁹ which reached similar conclusions about the locus of LTP.

How have the two Californian groups^{1,2} overcome these difficulties? Their chief success was to record from cells using low-resistance patch electrodes that allow an order-of-magnitude increase in signal-

- Jarrold, M.F., Ray, U. & Creagan, K.M. *J. chem. Phys.* **93**, 224–229 (1990).
- Andres, R.P. *et al. J. Mater. Res.* **4**, 704–736 (1989).
- Gupta, P., Mak, C.H., Coon, P.A. & George, S.M. *Phys. Rev.* **B40**, 7740–7749 (1989).
- Jarrold, M.F. in *Gas Phase Inorganic Chemistry* (ed. Russell, D.H.) 137–192 (Plenum, New York, 1989).
- Elkind, J.L., Alford, J.M., Weiss, F.D., Laaksonen, R.T. & Smalley, R.E. *J. chem. Phys.* **87**, 2397–2399 (1987).
- Maruyama, S., Anderson, L.R. & Smalley, R.E. *J. chem. Phys.* (in the press).
- Beck, S.M. & Andrews, J.M. *J. chem. Phys.* **91**, 4420–4421 (1989).
- Buffat, Ph. & Borel, J.-P. *Phys. Rev.* **A13**, 2287–2298 (1976).
- Knickerbein, M.B. & Yang, S. *J. chem. Phys.* **93**, 1476–1477 (1990).
- Kaxiras, E. *Phys. Rev. Lett.* **64**, 551–554 (1990).

to-noise ratio. They use two methods to reduce the number of activated synapses: minimal stimulation to activate at most a few afferent fibres; and impalement of a hippocampal presynaptic CA3 pyramidal cell projecting to an impaled postsynaptic CA1 cell to allow a single synapse to be studied (the success rate in such an endeavour is extremely low¹⁰ and Bekkers and Stevens resorted to dissociated hippocampal cultures for this type of experiment). Malinow and Tsien validated their approach by showing that it gives the expected answers in situations in which the answer is known. When transmission was enhanced by increasing extracellular calcium, there was a predicted increase in $1/CV$ (ref. 11), where CV is the coefficient of variation of minimal synaptic responses, a presynaptic measure depending only on N and P . Conversely, when a glutamate receptor antagonist was applied, leading to a decrease in synaptic transmission by a presumed reduction in a , there was no change in this measure.

In the case of experiments on hippocampal slices, both groups^{1,2} emphasize that their analysis is restricted to a composite measure of presynaptic function involving both N and P . Their results thus do not preclude a change in a , and hence some postsynaptic involvement. But in the experiments with cultured cells, Bekkers and Stevens found that the binomial distribution gave a better fit than the Poisson distribution, and showed that the change was in P rather than in N . This result is important because although a relatively implausible postsynaptic model² can be constructed to accommodate changes in N , it is difficult to see how this can be done for changes in P . Bekkers and Stevens can therefore make a slightly stronger claim for presynaptic suzerainty than Malinow and Tsien, although the fact that different statistical models apply in hippocampal slices and in cultured cells carries an implicit reminder that different mechanisms could apply in the two cases.

One surprising observation, however, applies to both preparations. Bekkers and Stevens report a substantial variance in the amplitude of individual quantal events (miniature excitatory synaptic currents), in contrast to the constancy of such events in other neurons where it has been examined⁹ — it is this additional variance that presumably obscures the appearance of peaks, corresponding to multiples of the quantal amplitude, in the amplitude distributions illustrated by Malinow and Tsien. In another paper¹², Bekkers *et al.* conclude that variability in miniature quantal events probably results from variation in vesicle size, and hence in the amount of transmitter released in each quantal event. But estimates of the number of glutamate molecules in a synaptic vesicle (about 3,000; ref. 11) are an order of magnitude greater than the estimates

of the number of kainate/quisqualate channels which need to open to provide the conductance change underlying a miniature quantal event. (In slices, the mean miniature quantal conductance is 0.8 nS (ref. 12). Assuming a kainate/quisqualate channel conductance of 10 pS, fewer than 100 channels open during a miniature quantal event.) Thus, even if cooperative binding of transmitter molecules is required for channel opening, a single vesicle probably contains enough transmitter to saturate subsynaptic receptors, so it is not clear how vesicle variability contributes to quantal variance.

Is this the end of the story? The history of the field suggests not. It remains to be seen whether quantal analysis of components mediated by receptor subtypes can reconcile the new results with the differential potentiation of NMDA and non-NMDA responses^{4,5}. Meanwhile, another report¹³ of quantal analysis of LTP in the Schaffer collateral pathway, which links CA3 cells to CA1 cells in the hippocampus, concludes that LTP is expressed as a change in quantal size without change in quantal content. Back on the fence, seats are filling up fast.

Finally, a presynaptic site for the

expression of LTP suggests that a signal molecule is required to carry information back across the synapse from the postsynaptic site of induction. One such candidate, arachidonic acid, has been identified¹⁴, and it is certain that the two new papers will stimulate the search for others. □

T. V. P. Bliss is at the National Institute for Medical Research, The Ridgeway, Mill Hill, London NW7 1AA, UK.

1. Malinow, R. & Tsien, R. *Nature* **346**, 177–180 (1990).
2. Bekkers, J. & Stevens, C. *Nature* **346**, 724–729 (1990).
3. Collingridge, G. L. & Bliss, T. V. P. *Trends Neurosci.* **10**, 988–990 (1987).
4. Lynch, G. & Baudry, M. *Science* **224**, 1057–1063 (1984).
5. Bliss, T. V. P. & Lynch, M. A. in *Long-Term Potentiation: From Biophysics to Behavior* (eds Landfield, P. W. & Deadwyler, S. A.) 3–72 (Alan R. Liss, New York, 1988).
6. Kauer, J. A., Malenka, R. C. & Nicoll, R. A. *Neuron* **1**, 911–917 (1988).
7. Mulier, D., Joly, M. & Lynch, G. *Science* **242**, 1694–1697 (1988).
8. Redman, S. *Physiol. Rev.* **70**, 165–198 (1990).
9. Voronin, L. L. *Neuroscience* **10**, 1051–1069 (1983).
10. Friedlander, M. J., Sayer, R. J. & Redman, S. J. *J. Neurosci.* **10**, 814–825 (1990).
11. Riveros, N., Fiedler, J., Lagos, C. & Orrego, F. *Brain Res.* **386**, 405–408 (1986).
12. Bekkers, J. M., Richerson, G. B. & Stevens, C. F. *Proc. natn. Acad. Sci. U.S.A.* **87**, 5359–5362 (1990).
13. Foster, T. C. & McNaughton, B. L. *Hippocampus* (in the press).
14. Williams, J. H., Errington, M. L., Lynch, M. A. & Bliss, T. V. P. *Nature* **341**, 739–742 (1989).

QUANTUM MECHANICS

A quantum time machine

Tony Sudbery

IDEAS from science fiction often find their way into serious theoretical physics, but one idea which has always seemed unlikely to cross the border between science and fiction is that of a time machine. Nevertheless, this has now happened with the publication in *Physical Review Letters* (64, 2965–2968; 1990) of a paper by Yakir Aharonov, Jeeva Anandan, Sandu Popescu and Lev Vaidman which contemplates the prospect of a “quantum time-translation machine”.

Time translation is a theoretical concept arising from physicists’ habit of treating time and space in the same way. A translation in space is the operation of moving objects in some direction through a given distance. Analogously, a time translation is the imagined operation of moving an event to a different position in time. Until the paper of Aharonov *et al.*, nobody would have supposed that such a translation could be anything other than a purely mental process.

The principle behind their time-translation machine is the fundamental feature of quantum mechanics termed superposition — the combining of different physical situations to give a new situation in which, by the indeterminism inherent in quantum mechanics, either of the original situations might manifest itself. Something of this

idea can be grasped by thinking of superposing photographic images, but this does not prepare one for the crucial point that the relative contributions of the two original situations to the superposition are measured by numbers which can be negative.

Part of the mystery of the principle of superposition lies in the fact that the things that are multiplied by (possibly negative) coefficients and then combined algebraically are not measurable quantities but are larger and more abstract: whole situations, histories, states of affairs. Aharonov *et al.* show that in certain conditions this abstract superposition can be reduced to the more familiar arithmetical addition of numerical quantities. They apply this to forces (technically, to hamiltonians) by considering a system in which the force can be set by some external control, and then arranging for a superposition of the situations arising from different settings. This, they show, comes down to simply taking a weighted average of the possible forces. But quantum mechanics, by allowing negative weights, leads to the possibility of a weighted average that lies far outside the range of the individual forces. So a number of small forces can be made to ‘average’ to a large force.

So far this describes nothing more con-

ceptually startling than a new form of amplification. What raises the possibility of time translation is the very direct relationship between the strength of a force (in the hamiltonian sense) and the length of time it takes to produce a given effect: doubling the hamiltonian has the same net effect as doubling the period of time for which it operates. The use of superposition suggested by Aharonov *et al.* can therefore be applied to periods of times instead of forces. If one could superpose a number of situations in which different periods of time have elapsed, one could, by suitably choosing the weights in the superpositions, obtain the effect of a lapse of time longer than in any of the original situations — thus taking the system into the future — or even of a negative time, so taking it into the past.

But how can one produce situations, all co-existing at *one* time, in which a force has been allowed to operate for *different* times? To achieve this, Aharonov *et al.* appeal to the general theory of relativity, according to which the time that elapses in a particular place, as measured over a given period of time by an observer in a different place, depends on the gravitational potential in the first place. Thus one could conceivably control the lapse of time by controlling the gravitational potential. For example, Aharonov *et al.* suggest that this can be done by enclosing the system in a massive spherical shell and then altering the radius of the shell.

Aharonov *et al.* have developed the first of these ideas into what they claim will be a practical method of amplification. But there are severe difficulties in the way of using the second idea to construct an actual time-translation machine. First, as Aharonov *et al.* acknowledge, the machine can in principle work only very rarely, as it relies on the random, indeterminate processes of quantum mechanics to produce the requisite state of superposition. Second, states of superposition are delicate affairs, requiring precise and intricate coherence between different parts of a system; they have never been produced in macroscopic apparatus such as Aharonov *et al.* describe. Finally, there are notoriously two ways in which quantum systems develop in time. There is the natural development which happens if the system is left undisturbed, and there are the quantum jumps which occur at a measurement, representing the inevitable disturbance caused by observation. The proposed machine takes the system into what would be the future (or the past) if only the first of these was operating; this is not the same as the actual future or past. So don't hold your breath waiting to meet H.G. Wells's Eloi and Morlocks. □

Tony Sudbery is in the Department of Mathematics, University of York, Heslington, York YO1 5DD, UK.

Messenger gets the green light

Robin Irvine

Now that inositol 1,4,5-trisphosphate (InsP₃) has been established as a second messenger in animal cells¹, relaying hormonal signals received at the cell surface to intracellular compartments, an exciting question pervades the world of plant physiology — do plants (especially higher plants) also use InsP₃ in this way? This is by no means a foregone conclusion, especially if one considers the *débat* of cyclic AMP, which in the late 1960s was assumed (on a mass of spurious flawed evidence) to be the mediator of many, even most, plant hormone actions. Two papers in this issue^{2,3} take a positive step towards placing InsP₃ in higher plants on a much firmer footing. They show that if chemically inactivated, or 'caged', InsP₃ is micro-injected into stomatal guard cells of two plants, *Commelina communis*² and *Vicia faba*³, then when the InsP₃ is liberated by a laser flash the intracellular calcium increases², the cells go through their physiological function of stomatal closure², and potassium channels (known to be controlled by Ca²⁺) are reversibly inactivated³. Gilroy *et al.*² also show that releasing 'caged' Ca²⁺ inside the cells in a similar way induces the same effects.

So, assuming that the same cells produce InsP₃ when exposed to a natural regulator such as abscissic acid, it is highly likely that InsP₃ is indeed an intracellular messenger controlling stomatal closure. We can probably assume that it acts on a specific receptor of the same ilk as that found in animal cells⁴, an assumption reinforced by the recent demonstration⁵ that vacuolar membranes from red beet have a Ca²⁺-carrying channel highly sensitive to micromolar quantities of InsP₃. Taken together^{2,3,5}, these data make the concept of InsP₃ as a second messenger in higher plants a comfortable idea to live with.

However, the stomatal response^{2,3} is a rapid one that occurs over only a few minutes — not exactly racing by mammalian standards admittedly, but when viewed against the life cycle of a bristlecone pine this is definitely the fast end of the timescale. These observations therefore add a new urgency to the question of whether plant hormone responses of the more conventional sort (hours to days, and upwards) also use the inositide signalling system. There, things are much less certain.

First, detecting a hormonally-induced response has been extremely difficult. To date, there is no unequivocal demonstration of increased InsP₃ formation caused by a natural stimulus in higher plants. Plants, unfortunately, show a much greater ingenuity than animals when it comes to a question of what to do with inositol⁶.

Feeding tritiated *myo*-inositol to plant tissues does indeed result in some radio-labelling of phosphatidylinositol 4,5-bisphosphate, PtdIns(4,5)P₂ (refs 7,8), the precursor of InsP₃, but it is present only in extremely small concentrations, and the analysis is complicated by the presence of other lipids that incorporate inositol, look like PtdIns(4,5)P₂ on thin-layer chromatograms, but are other things entirely⁹. (But note that PtdIns(4)P, the precursor of PtdIns(4,5)P₂, is the lipid which in cultured tomato cells turns over the fastest¹⁰.) The water-soluble components of [³H]*myo*-inositol-labelled tissues show a similar complexity¹¹, and many of the radioactive anionic compounds detected are not inositol phosphates at all. Against this alarming background of new knowledge, earlier claims that high concentrations of PtdIns(4,5)P₂ (similar to those in animals) exist in higher plant tissues, and that growth-promoting hormones stimulate the formation of InsP₃ (ref. 11 and references therein) must be viewed with some caution. In short, although higher plants do have the necessary enzymatic machinery to make and respond to InsP₃, we still have little idea what they use it for (apart from controlling stomatal closure).

In animal cells, all but the most acute responses (less than a minute) mediated by inositides involve second messengers in addition to InsP₃. So, in view of the fact that we have evidence for the involvement of InsP₃ only in rapid plant responses^{2,3,12}, it is relevant to ask whether plants also use these other messengers. For example, an animal cell needs to stimulate the entry of Ca²⁺ from the extracellular milieu to keep its Ca²⁺ concentration high for anything longer than necessary for a transient response.

As the mist surrounding the probable role of inositol 1,3,4,5-tetrakisphosphate (InsP₄) in controlling the entry of Ca²⁺ in animal cells begins to clear a little^{1,13}, it is pertinent to ask whether plants can also make this compound. *Chlamydomonas eugametos* has a very active and specific InsP₃-3-kinase (L. A. Stephens, A. Letcher, A. Musgrave and myself, unpublished data), and cell-free extracts of pea roots can phosphorylate InsP₃ into a substance which is probably InsP₄ (J. Chattaway and B. Drøbak, personal communication). But *Chlamydomonas* is a lower plant containing amounts of PtdIns(4,5)P₂ similar to those found in animal cells⁷, and the putative InsP₃-kinase activity in pea roots is as yet incompletely characterized, so this question is still open. Although Ca²⁺ influx has been demonstrated in stomatal guard cells, evidence for hormonal stimulation remains elusive¹⁴. It is also true that

the vacuole in the average vacuolated cell probably contains enough Ca^{2+} to keep the cytoplasm supplied for years, so if that organelle is indeed the target for InsP_3 (refs 5,15), then one could certainly question whether higher-plant cells would always need a Ca^{2+} -entry mechanism.

The role of diacylglycerol as a second messenger in higher plants (to stimulate protein kinase C) is similarly unclear. There are no convincing reports of a diacylglycerol-stimulated protein kinase, and although a search of two higher-plant genomes for something resembling protein kinase C produced homologues¹⁶, they are considered markedly different and are much more likely to be regulated by something other than diacylglycerol: lysophospholipids are one possibility¹⁷.

Thus, out of the three second messengers which in animals cells mediate the 'classic' phosphoinositide response — InsP_3 , InsP_4 and diacylglycerol — the role of the last two in plants is uncertain.

In conclusion, even if we take it that higher plants do indeed use InsP_3 to modulate rapid responses, we remain in frustrating ignorance about its place, and the place of the phosphoinositide cycle in general, in the physiology of higher plants. So the short answer to the question of whether higher plants use InsP_3 as a second messenger seems to be yes — but for how long? □

Robin Irvine is at the AFRC Institute of Animal Physiology and Genetics Research, Babraham, Cambridge CB2 4AT, UK.

- Berridge, M. & Irvine, R. *Nature* **341**, 197–205 (1989).
- Gilroy, S., Read, N.D. & Trewavas, A.J. *Nature* **346**, 769–771 (1990).
- Blatt, M.F., Thiel, G. & Trentham, D.R. *Nature* **346**, 766–769 (1990).
- Ferris, C.D., Haganis, R.L., Supattapone, S. & Snyder, S.H. *Nature* **342**, 87–89 (1989).
- Alexandre, J., Lassalles, J.F. & Kado, R.D. *Nature* **343**, 567–570 (1990).
- Loewus, R.A. & Loewus, M.W. *A. Rev. Pl. Physiol.* **34**, 137–161 (1983).
- Irvine, R.F., Letcher, A.J., Drabak, B.K., Dawson, A.P. & Musgrave, A. *Pl. Physiol.* **89**, 888–892 (1989).
- Coté, G.G. et al. *Pl. Physiol.* **90**, 1422–1428 (1989).
- Drabak, R.K. et al. *Pl. Physiol.* **87**, 217–222 (1988).
- Coté, G.G., Morse, M.J., Satter, R.L. & Crain, R.C. *Pl. Cell Reps.* **6**, 352–355 (1987).
- Ettlinger, C. & Lehle, L. *Nature* **331**, 176–178 (1988).
- Morse, M.J., Crain, R.C. & Satter, R.L. *Proc. natn. Acad. Sci. U.S.A.* **84**, 7075–7078 (1987).
- Irvine, R.F. *FEBS Letts* **263**, 5–9 (1990).
- MacRobbie, E.A.C. *Planta* **178**, 231–241 (1989).
- Schumaker, K.S. & Sze, H. *J. biol. Chem.* **262**, 3944–3946 (1987).
- Lawton, M.A., Yamamoto, R.T., Hanks, S.K. & Lamb, C.J. *Proc. natn. Acad. Sci. U.S.A.* **86**, 3140–3144 (1989).
- Martiny-Baton, G. & Scherer, G.F.E. *J. biol. Chem.* **264**, 18052–18059 (1989).

MANTLE CONVECTION

Deep causes of hotspots

Marcia McNutt

OF the three principal types of volcanoes on the Earth's surface, those at mid-ocean ridges and island arcs are relatively well understood as melting processes triggered by the relative motion of tectonic plates in divergent and convergent zones, respectively. Hotspot volcanism, which produces linear chains of volcanoes (the Hawaiian Islands, for example) as the plates drift over melt 'plumes' fixed with respect to the deep mantle, is not explained by the theory of plate tectonics and thus represents one of the few geological phenomena on the Earth's surface that directly provide information on processes in the convecting mantle. In a new compilation of a mass of geophysical data¹, Sleep has deduced the flux of material driving the activity of 37 hotspots, yielding new insights into their global effect.

Sleep's method for calculating the characteristics of each plume depends on its location. For hotspots that have erupted on thickened lithosphere away from mid-ocean ridges, he calculates the total amount of buoyant plume material from the cross-sectional area of the swell that develops in the topography and geoid (height of the Earth's gravitational equipotential surface) as hot material is injected into a thin low-viscosity channel at the base of the plume. The mass

flux from the plume is independently estimated from the shape of the swell upstream from the direction of plate motion, assuming that it represents the stagnation point of radial flow from the

plume interacting with the mantle flow induced by drag of the lithospheric plate. The excess temperature of the plume can then be calculated given its buoyancy and mass fluxes.

For hotspots near mid-ocean ridges, all of the plume flux is assumed to rise to sufficiently shallow depths to cause melting. The mass flux is thus derived from the amount of excess volcanism, as revealed by the topography and crustal thickness.

From the excess temperature of the plumes relative to the surrounding mantle, typically 200 °C, Sleep tries to constrain values of the temperature contrast across the core-mantle boundary, from which he assumes the plume heat flux emanates. He argues that the usual lower bound for this contrast, 1,000 °C (as derived from experimental melting curves for iron²), is not possible given his geophysical data. Furthermore, using his estimate that hotspots cool the Earth's core at a rate of 50 °C per thousand million years, Sleep suggests that the temperature contrast at the core-mantle boundary has been sufficient to drive plume-type activity only over the past 1,300 million years. But both conclusions depend critically on the character of the core-mantle boundary. If it comprises both a thermal and a chemical boundary layer, the latter of which seems necessary to reconcile discrepancies between seismic and gravitational studies of the core-mantle topography, Sleep's temperature contrast could apply solely to that across the thermal boundary layer at the base of the convecting mantle from which he assumes the plumes arise. The deeper chemical boundary layer serves to insulate

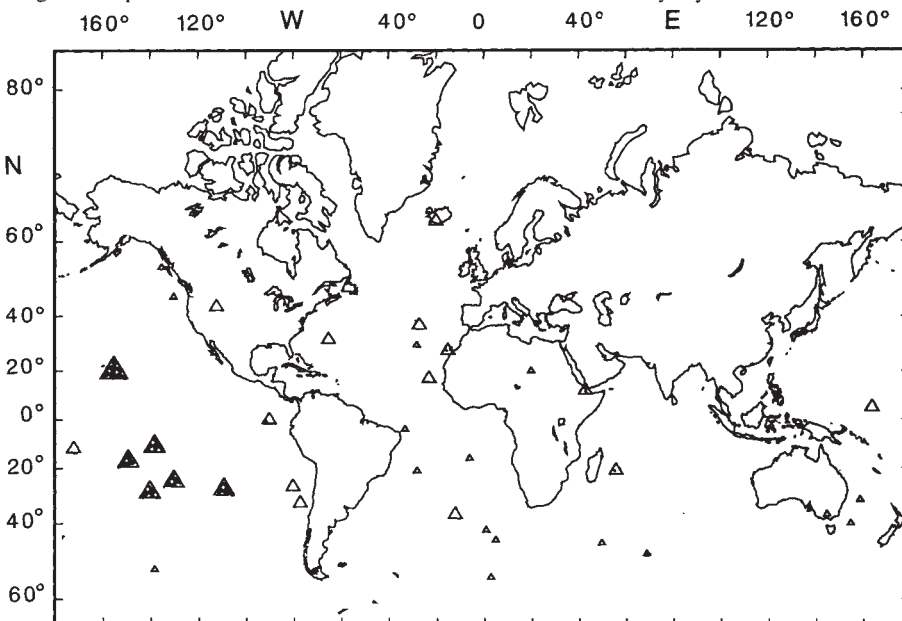


FIG. 1 Map showing the location of hotspots for which the buoyancy flux has been calculated¹. Buoyancy flux is defined as $\rho_m \alpha \Delta T Q$, in which ρ_m is the density of the mantle, α is the coefficient of thermal expansion, ΔT is the average excess temperature of the plume (about 200 °C), and Q is the volume of hotspot material supplied per unit time. Symbol size proportional to the relative magnitude of the flux. Shaded symbols, Hawaiian and Superswell hotspots.

the core from the mantle, thus allowing the core to be much hotter, in agreement with the iron-melting data.

Sleep also concludes that most of the heat flux from mantle plumes is emplaced into the lower oceanic lithosphere, which is subsequently subducted. Thus the surprising result is that mantle plumes, via a circuitous route through the lithosphere, effect a transfer of heat from the Earth's core to the mantle.

Viewed in another way, Sleep's flux estimates can be used to settle the decade-old controversy of whether the non-random distribution of hotspots on the Earth's surface (Fig. 1) represents the fundamental structure of plume-type upwelling in the mantle³, or the ease with which the overlying lithosphere can be penetrated by melt⁴. Proponents of the first view point have demonstrated that most hotspots are located beneath long-wavelength highs of the geoid, whereas plume-free regions correlate with long-wavelength geoid lows. This global pattern of geoid highs and lows is thought to reflect broad upwelling and downwelling cells in the convecting mantle, which presumably encourage and discourage, respectively, the ascent of plumes from the deep mantle.

Proponents of the lithospheric-control viewpoint have argued that the distribution of hotspots is controlled by the ease with which rising melt from the mantle can penetrate the overlying, cold, lithospheric lid, and thus provides little information on mantle circulation. Because the thickness of the lithosphere increases as the square root of plate age, older lithosphere should be less vulnerable to hotspot volcanism than younger lithosphere. Furthermore, a plate that drifts slowly with respect to the plume source should be more easily penetrated than one that quickly sweeps past it, allowing little time for transfer of heat and melt. But judging solely on the distribution of hotspots, it is difficult to choose between the mantle- and lithospheric-control hypotheses, because there are exceptions to the predictions of both models (hotspots located on geoid lows or erupting on old, rapidly drifting plates). With Sleep's quantification of hotspot flux, however, it is possible to assess whether the exceptions to each rule are significant in terms of the mass and heat carried by hotspots.

Figure 2 shows histograms of hotspot flux as a function of various parameters likely to represent the lithosphere's ability to act as a barrier of hotspot penetration. Because Sleep's hotspot compilation omits some examples on continental lithosphere (on account of difficulties in analysing the relevant data), I have concentrated on the flux from just the oceanic hotspots and normalized the flux with respect to the area of oceanic

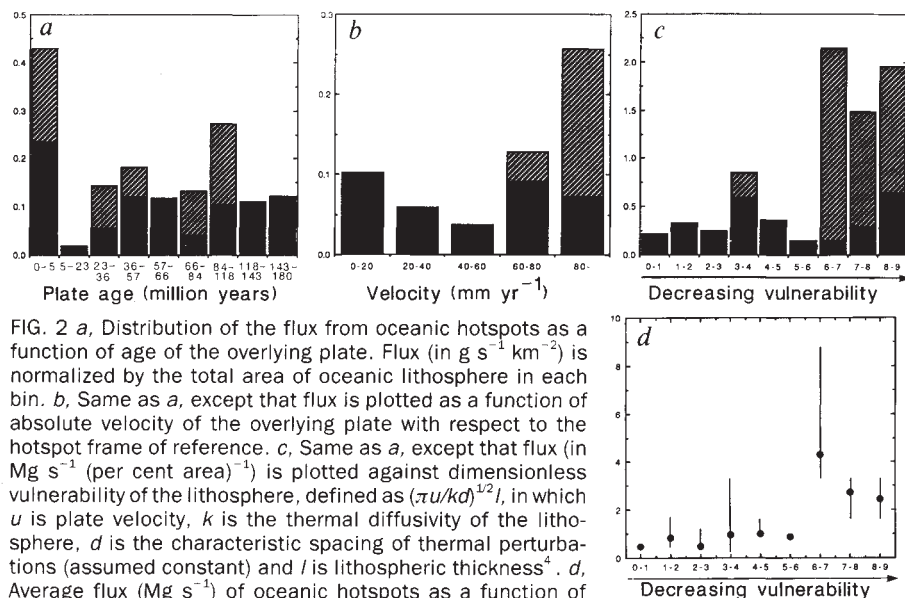


FIG. 2 *a*, Distribution of the flux from oceanic hotspots as a function of age of the overlying plate. Flux (in $\text{g s}^{-1} \text{km}^{-2}$) is normalized by the total area of oceanic lithosphere in each bin. *b*, Same as *a*, except that flux is plotted as a function of absolute velocity of the overlying plate with respect to the hotspot frame of reference. *c*, Same as *a*, except that flux (in $\text{Mg s}^{-1} (\text{per cent area})^{-1}$) is plotted against dimensionless vulnerability of the lithosphere, defined as $(\pi u/kd)^{1/2}$, in which u is plate velocity, k is the thermal diffusivity of the lithosphere, d is the characteristic spacing of thermal perturbations (assumed constant) and l is lithospheric thickness⁴. *d*, Average flux (Mg s^{-1}) of oceanic hotspots as a function of vulnerability. Vertical bars show the total range of values in each vulnerability category. (Shaded bars, flux from Hawaiian and Superswell hotspots; black, flux from all other hotspots.)

lithosphere represented in each histogram category. The flux from plumes is discharged without regard to the age of the overlying plate (Fig. 2*a*), except for the two youngest age bins which display an excess of flux near spreading ridges (less than 5 million years old) and a compensating deficit on relatively young lithosphere, found slightly further from spreading centres. As Sleep suggests, this could be either the result of the capture of a nearby ridge by a plume or the sign of entrainment of the flux from a nearby plume into the rising flux at the ridge.

The flux liberated as a function of velocity of the plate also shows no tendency for rapid plate motion to inhibit the flux from hotspots (Fig. 2*b*). Plotting the flux as a function of the 'lithospheric vulnerability parameter', which increases linearly with plate thickness and as the square root of plate velocity⁴, again shows that most of the hotspot flux penetrates lithosphere that is predicted to have low vulnerability for volcanism (Fig. 2*c*). The one parameter that does appear to be sensitive to plate vulnerability is the average flux of the hotspots (Fig. 2*d*). Feeble hotspots simply do not occur on lithosphere with low vulnerability (large thickness and/or rapid drift velocity). On the other hand, 85 per cent of the hotspot flux in Sleep's compilation is correlated with positive values of the geoid when referenced to the best-fitting ellipsoid (but 70 per cent when referenced to the hydrostatic figure).

In fact, the impression from Fig. 2*c* is that lithosphere of low-vulnerability, by virtue of its rapid motion (Fig. 2*b*) rather than old age (Fig. 2*a*), is, ironically, the site for most of the global flux of hotspot volcanism. This conclusion is entirely dictated by the large flux from Hawaii and the Superswell⁵ hotspots of French Polynesia, which together account for 45

per cent of the global flux. This correspondence between rapid plate motion and large plume flux holds not only for the present, but also for a previous period of rapid plate motion 70–100 million years ago when massive volcanism from Superswell hotspots produced the Darwin Rise seamount chains now located in the northwest Pacific⁶. If indeed mantle circulation (as indicated by the global geoid pattern) controls the distribution of plume flux, then the large discharge of plume flux through rapidly drifting plates with low vulnerability may point to a connection between rapid plate motion and mantle upwelling.

This link between hotspots and plate speed is quite distinct from earlier suggestions⁷ that plumes must be part of a broad flow of mantle driven by plate motions. This idea arose because hotspots cluster near mid-ocean ridges and avoid subduction zones. I now doubt the evidence used to support this new theory. The analysis¹ demonstrates that the ridge-related entrainment of flux is only a local phenomena (Fig. 2*a*; 23 million years in time is equivalent to 1,000 km of movement), and plumes near subduction zones might easily be mistaken for island-arc type volcanism. Indeed, some arc systems show isotopic and trace element signatures derived from a plume source⁸. □

Marcia McNutt is in the Department of Earth Sciences, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA.

1. Sleep, N.H. *J. geophys. Res.* **95**, 6715–6736 (1990).
2. Williams, Q. *et al. Science* **236**, 181–182 (1987).
3. Vogt, P.R. *J. geophys. Res.* **86**, 950–960 (1981).
4. Pollack, H.N., Gass, I.G., Thorpe, R.S. & Chapman, D.S. *J. geophys. Res.* **86**, 961–966 (1981).
5. McNutt, M. & Judge, A. *Science* **248**, 969–975 (1990).
6. McNutt, M.K., Winterer, E.L., Sager, W.W., Natland, J.H. & Ito, G. *Geophys. Res. Lett.* (in the press).
7. Houseman, G. *Nature* **340**, 263–264 (1989).
8. Morris, J.D. & Hart, S.R. *Geochim. Cosmochim. Acta* **47**, 2015–2030 (1983).

Neurofibromatosis gene cloned

Bruce Ponder

ANOTHER high-profile human disease has fallen to the strategy of genetic linkage followed by detailed mapping and cloning of the gene¹⁻³. This time it is the gene for type-1 (Von Recklinghausen) neurofibromatosis (NF-1), a disease that affects an estimated 1 in 3,500 of the general population. This type of neurofibromatosis is the autosomal dominant disorder popularly (though erroneously) associated with Joseph Merrick, the 'Elephant man'. Sequence homology suggests that the *NF-1* gene product shares similarities with the mammalian GAP gene⁴ (see Wigler's News and Views article on page 696 of this issue). The widespread effects of the *NF-1* mutation suggest that its biology will be interesting. Moreover, the gene has already provided a surprise by having embedded within its introns several other genes, transcribed in the opposite orientation, a feature previously described in man only in the factor VIII gene⁵.

NF-1 primarily affects tissues derived from the neural crest (see table)^{6,7}. Vari-

bility is a prominent feature. The biology is a puzzle — it is not even known which cell types are primarily affected. For example, in the common skin neurofibromas (see figure, overleaf) the histological appearances suggest a disordered interaction between axons and the surrounding Schwann cells. If so, the primary defect could lie in either cell type, or in both. The origin of the tumours is not understood either. Both histology and X-chromosome inactivation studies indicate that neurofibromas are polyclonal, suggesting a mechanism perhaps involving local cellular interaction: but the presence of a primary neoplastic clone within the mixed population cannot be excluded.

The *NF-1* gene was mapped to the long arm of chromosome 17 in 1987. Later, two chromosome translocations involving 17 (t(1;17) and t(17;22)) were found in two NF-1 individuals: pulsed-field gel electrophoresis showed that the translocation breakpoints lie only about 60 kilobases apart, and so it seemed the *NF-1* gene had

been tightly localized. It took much longer than expected to find the gene, because although several candidates expressed in nervous tissue were quickly found, they were all rather small and none could be shown to be mutated in NF-1 patients. This is not what was expected: the very high mutation rate in NF-1 (about 1 in 10⁴; almost 50 per cent of patients have new mutations) suggested either that the gene would be very large, or that there would be frequent DNA rearrangements indicating the location of the gene.

The finding of the present candidate *NF-1* gene illustrates the value of a detailed DNA map of the candidate region, but also the importance of the single rare case with an informative mutation. The Michigan group¹ screened complementary DNA (cDNA) libraries with a conserved sequence found by chromosome jumping, and also with an entire yeast artificial chromosome (YAC) clone spanning the translocation region. The group found two clones covering a sequence of 2,012 base pairs with a single open reading frame containing six exons, and spanning at least 33 kilobases of genomic DNA. Northern blots indicate hybridization with messenger RNA (mRNA) of roughly 13 kilobases. The evidence that this clone is part of the *NF-1* gene is first, that studies in human-rodent hybrids containing the isolated (1;17) and (17;22) translocations show that both the gene sequence and its transcript are disrupted by the (17;22) breakpoint, and second, that one NF-1 patient out of 35 studied had a novel 500-base-pair insertion within or close to an exon of the candidate sequence, absent from his parents' DNA.

The Salt Lake City group^{2,3} found that 3 out of 54 NF-1 patients had germ-line deletions in the translocation region. The smallest deletion, of 11 kilobases, did not include any of the previous gene candidates but did include a conserved sequence spanning the (17;22) breakpoint. Using the conserved sequence to screen cDNA libraries, this group first obtained 4 kilobases of overlapping DNA sequence that contained a single open reading frame of about 3,300 base pairs, oriented 5' to 3' in the centromere-to-telomere direction. This corresponds to nine exons spaced over a 100-kilobase genomic region extending 65 kilobases from the (17;22) breakpoint in the direction of the telomere.

The evidence that this was part of the *NF-1* gene was first that the 11-kilobase *NF-1* genomic deletion included this gene but not the other candidates; second that the gene was interrupted by the (17;22) breakpoint; and finally, a search for alterations in DNA sequence in exons 4-9 of the candidate gene using single-strand conformation polymorphism⁸ revealed six variant alleles in 72 NF-1 individuals and none in 65 controls (although two other

SOME FEATURES OF NF-1

	Approximate % of gene carriers affected	Probable cell of origin	Comments
Diagnostic features			
Café-au-lait spots (more than 5)	>95	Melanocyte	1 cm or larger pigmented skin patches.
Skin neurofibromas (up to 500 or more)	>90	? Schwann cell (Multiple cell types)	Develop at puberty. Non-malignant. Contain Schwann cells, axons, fibroblasts, mast cells, perineural cells and blood vessels.
Lisch nodules	>90	?	Nodular developmental abnormalities of iris.
Less common problems			
Learning difficulties	30	?	
Short stature	15	?	less than third centile.
Large head	25	?	more than 97th centile.
Skeletal abnormalities:			
Scoliosis	30	?	
Pseudoarthrosis	3	?	Failure of development of one segment of a long bone (usually tibia).
Epileptic fits	<5	?	
Tumours			
Plexiform neurofibromas	20	? Schwann cell (multicellular like skin neurofibromas)	Can cause severe deformity; may become malignant; May be associated with local hypertrophy.
Neurofibromas of spinal and peripheral nerves*	5	As above	
Schwannomas of peripheral nerves	<5	Schwann cell	Occasionally become malignant.
CNS tumours (glioma, ependymoma, occasional meningioma)	3	Glia, ependyma, arachnoid	Optic nerve gliomas relatively frequent often nonprogressive.
Phaeochromocytoma	1	Chromaffin cell of adrenal medulla	Usually benign.
Neurofibrosarcomas; malignant Schwannoma	2	? Schwann cell	Malignant: usually develop from plexiform NF or Schwannoma.

* These tumours are prominent in some families with a variant form of *NF-1* which maps to the *NF-1* locus.

IMAGE UNAVAILABLE FOR COPYRIGHT REASONS

Neurofibromas in NF-1. Left, typical appearance of skin neurofibromas; right, plexiform neurofibroma is present in the lower left leg. Note the local overgrowth of the tissue; the leg is also longer. Plexiform neurofibromas can also be associated with thickening, pigmentation and excessive hair growth of the overlying skin. Photographs from ref. 8.

non-NF-1 individuals did have variants). The candidate *NF-1* gene sequences identified by each group overlap. Subsequently, the Salt Lake City group obtained sufficient additional sequence information to reveal significant homologies between the NF-1 protein and mammalian GAP (ref. 4).

The gene is expressed in a wide variety of tissues, including many not of neural-crest origin: the specificity of the NF-1 mutation has therefore to be explained, but might involve tissue-specific modification or interaction of the *NF-1* gene product with other proteins.

Each of the NF-1 mutations described so far is expected to result in inactivation of the gene. Loss of the remaining allele by somatic mutation, as in retinoblastoma (*Rb*) is a possible mechanism of tumour formation; but it seems less plausible to account for more general manifestations such as short stature or learning difficulties. Dosage effects from loss of one copy seem more likely. More complicated (and speculative) schemes involve possible effects on the activity of the *NF-1* gene of transcription of the genes embedded in reverse orientation within its introns, either by steric hindrance or formation of RNA duplexes with the *NF-1* gene product, which would be degraded. It is not even clear whether the *NF-1* gene acts as an *Rb*-like tumour suppressor gene in relation to tumour formation. The predicted allele losses in the *NF-1* region in tumours have not been reported so far; but there could be selection for a gene near *NF-1* being retained in the tumours, explaining why the large-scale chromosomal events detectable by the allele loss assay¹⁰ might be rare. Detailed analysis of both alleles of the *NF-1* gene and of transcripts in populations of cells purified from tumours will be needed to resolve this question.

The practical, clinical consequences of the cloning are uncertain. Because of the high mutation rate, many pedigrees contain only one NF-1 individual, so prenatal

diagnosis using linked markers is impossible. Identification of the mutations will allow prenatal diagnosis, but the practicalities of this will depend on whether most mutations occur in a limited region of the gene. So far, examination of about 10 per cent of the coding sequence in 72 individuals has revealed 6 mutations, spread over 4 exons³.

Because many cases of NF-1 are so mild, for most people the most important prediction is likely to be not the inheritance but rather the severity of the disease. Some uncommon variants of NF do seem to run true in families⁶; for these, correlation with the type of germ-line mutation seems likely (and could be informative in terms of gene function). Most often, however, it seems that neither the individual features of NF-1 listed in the table, nor overall severity, breed true; the relative contribution to phenotype of germ-line mutation, modifying genes, environment and chance is still unclear. It will be interesting to see whether correlations can be established between types of germ-line mutation of the *NF-1* gene itself or the genes embedded within it and the presence of specific NF-1 phenotypes in families; and whether elucidation of the *NF-1* gene product and its action will provide clues to factors which modify its effects. □

Bruce Ponder is at the Cancer Research Campaign Human Cancer Genetics Research Group, Department of Pathology, University of Cambridge, Tennis Court Road, Cambridge CB2 1QP, UK.

1. Wallace, M.R. *et al.* *Science* **249**, 181–186 (1990).
2. Viskochil, D. *et al.* *Cell* **62**, 187–192 (1990).
3. Cawthorn, R.M. *et al.* *Cell* **62**, 193–201 (1990).
4. Xu, G. *et al.* *Cell* **62**, 599–608 (1990).
5. Levinson, B., Kenwick, S., Lakich, D., Hammonds, G., Jr & Gitschier, J. *Genomics* **7**, 1–11 (1990).
6. Riccardi, V.M. & Eichner, J.E. *Neurofibromatosis* (Johns Hopkins University Press, Baltimore and London, 1986).
7. Huson, S.M., Compston, D.A.S., Clark, P. & Harper, P.S. *J. med. Genet.* **26**, 704–711 (1989).
8. Huson, S.M., Harper, P.S. & Compston, D.A.S. *Brain* **111**, 1355–1381 (1988).
9. Orita, M., Iwahana, H., Kanazawa, H., Hayashi, K. & Sakiya, T. *Proc. natn. Acad. Sci. U.S.A.* **86**, 2766–2770 (1989).
10. Ponder, B. *Nature* **335**, 400–402 (1988).

High charges

LAST week Daedalus presented his 'ionic windmill'. A big wire grid with many discharge points emits electrons into the air, which form molecular ions with the air. These are blown downwind to a collector grid, whose return current to the emitter captures much of the wind's energy. He now plans to run the whole scheme in reverse, to create an ionic helicopter.

This elegant craft simply consists of two parallel horizontal wire meshes. The upper, negative mesh will be studded with downward-pointing sharp points coated with material of low work function. The ion emission from these points will be attracted towards the lower, positive mesh. Ions diffuse only slowly in air, so their downward flow will entrain a strong down-draught: the well-known 'electric-wind' phenomenon. The voltage to drive the craft will come via suitable inverter electronics from an array of solar panels mounted above the top grid: on edge, so as not to impede the downward airflow, and spaced so as not to shadow each other.

A structure of two open meshes braced to rigidity by insulated struts between them, and powered by solar panels, is already lighter and stronger than any other air-frame. The 'ionicopter' should be wonderfully efficient as well. To minimize the kinetic energy wasted by its down draught, a helicopter must push down the maximum volume of air at the minimum velocity. A wide wire-grid structure can capture far more air than any rotor, and can accelerate it gently and silently without wasteful turbulence.

The ionicopter will make an ideal aerospace vehicle. At take-off, a small outward-facing corona-discharge point will eject positive ions till the whole craft is negative enough to initiate downward electric-wind discharge from the upper grid. The potential difference between the grids will then be increased until the down-draught lifts the vehicle into the sky. Thereafter, the higher it goes, the better. Its Sun power becomes more reliable, the solar cells gain efficiency in the cold upper air, the electric discharge becomes more copious at low pressures, and the velocity of the down-draught rises to compensate for its falling density. Unlike any winged aircraft, it can soar almost into space. Given batteries to store surplus solar energy for use at night, it could stay up indefinitely.

Daedalus's ionicopter will compete strongly with satellites as a research vehicle, a TV relay and photoreconnaissance platform. At night, it will be visible from the ground by its auroral glow-discharges in the low pressure of aerospace. The ozone they generate will go to replenish the halocarbon-sabotaged ozone layer.

David Jones

Whale migration record

SIR—We recently confirmed a new long-distance migration record for mammals (apart from *Homo sapiens*). We found that a humpback whale (*Megaptera novaeangliae*) has migrated from the Antarctic peninsula region (64° 20'S; 62° 27'W) to Colombia (2° 57'N; 78° 12'W). The shortest swimming distance between these locations is more than 8,334 kilometres. We identified the whale at both locations using photographs of the distinctive ventral pigmentation on the tail flukes¹. The whale was seen in Antarctica on 19 April and in Colombia on 28 August 1986. This is the first time that a humpback whale has been shown to cross the equator, and the first time an Antarctic humpback whale has been documented in South American waters. This finding has interesting implications for the possibility of migration between hemispheres.

The match of pigmentation was found by comparing a catalogue of 32 humpbacks identified from the Antarctic peninsula² and a collection of 86 identified off the coast of Colombia. The whale swam from Antarctica to Colombia in less than 5 months. Following the typical humpback migration strategy of spending summer in high latitudes and winter in low latitudes³, this whale probably belongs to a population that feeds in Antarctica and breeds off the coast of Colombia.

It is also possible that Antarctic peninsula humpback whales migrate up the eastern side of South America to the coast of Brazil, where there is another breeding ground. The swimming distance from the Antarctic peninsula to this area is shorter by more than 1,800 kilometres than to the Colombian site. This raises the interesting possibility that the Antarctic peninsula feeding grounds may be an area from which humpback whales can cross between the South Pacific Ocean and the South Atlantic Ocean. It is of interest to note that only 1,000 kilometres further north of our Colombian study site there is a site used during the breeding season by humpbacks that feed in the Northern Hemisphere (G. H. Steiger, J. Calambokidis and R. Sears, unpublished data).

Despite opposite migration schedules, it might be possible for an occasional individual from the Southern Hemisphere to breed with members of this Northern Hemisphere population. If this also occurred on the east side of South America, it would offer a mechanism for Baker's supposition⁴ that the relatively small genetic distance between the North Pacific and North Atlantic populations could be the result of periodic gene flow between oceans. Although previous research clearly demonstrates that most humpbacks probably do not cross the equator or move between oceans, the

possibility that individuals occasionally might do so has important implications for gene flow, population differentiation and population management, and deserves further study.

GREGORY S. STONE

NOAA, Undersea Research, R/OR2,
1335 East West High Way,
Silver Spring,
Maryland 20910, USA

LILIAN FLÓREZ-GONZALEZ

Fundacion Yubarta,
A.A. 33141 Cali, Valle,
Colombia

STEVEN KATONA

College of the Atlantic,
Bar Harbor,
Maine 04609, USA

1. Katona, S.K. & Whitehead, H. *Polar Rec.* **20**, 439–444 (1981).
2. Stone, G.S. & Hamner, W. *Polar Rec.* **24**, 15–20 (1988).
3. Mackintosh, N.A. *The Stocks of Whales* (Fishing News, London, 1965).
4. Baker, C.S. *et al. Nature* **344**, 238–240 (1990).

Rotating gyros and gravity

SIR—The report by Hayasaka and Takeuchi¹ that the gravitational acceleration of a rotating mass depends on both the magnitude and vectorial direction of the angular momentum suggests several considerations. It is well known^{2,3} that the motion of an extended body in the Einstein theory of gravitation may depend on spin and angular momentum: the free motion of an extended mass endowed with angular momentum need not be geodesic. There are several theories of gravitation other than Einstein's in which such effects are definitely predicted. Accordingly, excitement at the report of this effect should focus on its size and asymmetry and implications for Einstein's theory of gravitation rather than on the question of whether it exists.

The possibility that angular momentum may influence the gravitational acceleration of a rotating mass was perhaps first noted in 1912 by Nordström⁴ in a paper in which he tried to derive a Lorentz invariant theory of gravitation from a scalar potential. Nordström quoted a communication from Einstein, ending: "A simple example shows that in such a theory a rotating object in a gravitational field will have a smaller acceleration than one which is not rotating." Einstein's arguments were never published, and he never subsequently returned in print to the subject.

Explicit calculations of the effect have been published⁵, however. An analogous, extremely elementary calculation for the Einstein equations shows the effect, in contradiction to the results of Papapetrou

and Synge, to be (precisely) absent for the case of a small object with the angular momentum vector parallel to the gravitational field.

For the Nordström theory and newtonian theory modified to be compatible with special relativity, the gravitational acceleration of a rotating object is independent of both the mass of the object and the orientation of the angular momentum vector to the field direction. Rather, it depends on (roughly) the average tangential speed in the horizontal plane, when reduction of the acceleration of gravity is of the second order in v/c or about 1 part in 10^{10} . Qualitatively, it seems to be simply related to the fact that acceleration in special relativity depends, *inter alia*, on components of the velocity transverse to the acceleration. Similar effects are suggested in many other theories of gravitation.

The meticulous nature of the report of Hayasaka and Takeuchi allows for further experiments by others. It is to be hoped that the arrangement of equipment in some of these tests would measure the dependence of the effect on the magnitude of the polar angle of the angular momentum vector with respect to the vertical.

ALEX HARVEY

Department of Physics,
City University of New York,
Flushing, New York 11367-0904, USA

1. Hayasaka, H. & Takeuchi, S. *Phys. Rev. Lett.* **63**, 2701–2704 (1989).
2. Papapetrou, A. *Proc. R. Soc. A* **64**, 57–75; 302–310 (1951); *A* **209**, 248–258 (1951).
3. Synge, J.L. *Relativity: The General Theory* (Interscience, New York, 1960).
4. Nordström, G. *Physik Zeitschr.* **XIII**, 1126–1129 (1912).
5. Harvey, A. *Ann. Phys.* **29**, 383–390 (1964).

Sunspot cycles and climate

SIR—Geller¹ in News and Views discussed the implications of a report by James and James² in which an oscillation of about a decade long appeared in power spectra of certain output from a model of the global troposphere. Because the oscillation was generated internally — that is, within the atmospheric model, both authors concluded that "there is no reason to appeal to external forcings to account for ultra-low-frequency atmospheric variability". This statement is obviously based on the numerous claims made to attribute approximately decadal cycles in climate time-series to forcing by the nominal 11-year sunspot cycle, and that so far, no generally accepted and quantitatively based physical mechanism has been provided to explain a connection between solar and atmospheric variability. But if the new general circulation model simulations are correct, they could be interpreted as indicating that the search³ for a link between the sunspot cycle and weather

should actually be continued.

An important step in studying signals in time-series, once a signal of interest has been identified in the power spectrum, is to generate the waveform in the time domain. This has already been done¹ for the nominal 11-year wave occurring in many surface climate time-series. A similar waveform occurs in a 250-year snow accumulation record from a northern mid-tropospheric site⁵. These waveforms, particularly the latter, exhibit pronounced amplitude modulation and phase shifts with respect to the sunspot cycle, which itself exhibits amplitude as well as frequency modulation. The waveforms derived from climate data may be approximately in phase with the solar cycle for a certain number of cycles only, before passing through a low-amplitude disordered sequence followed by a further group of clear cycles where the phase relationship may be retained or it may switch 180° (ref. 4). With the exception of the latter effect, these observations may be substantially explained by the interaction of two atmospheric waveforms differing in period by about 1 to 2 years⁶. Thus, the solar cycle added to an approximately 10-year wave could produce the observed waveforms.

G. HOLDSWORTH

National Hydrology Research Institute,
Saskatoon, Saskatchewan,
Canada S7N 3H5

1. Geller, M.A. *Nature* **342**, 15–16 (1989).
2. James, I.N. & James, P.M. *Nature* **342**, 53–55 (1989).
3. Tinsley, B.A. et al. *J. geophys. Res.* **94**, 14783–14792 (1989).
4. Currie, R.G. & O'Brien, D.P. *J. Clim.* **8**, 255–281 (1988).
5. Holdsworth, G. et al. *Publ.* **183**, 71–79 (Int. Ass. hydrol. Sci., 1989).
6. Castagnoli, C. et al. *J. geophys. Res.* **94**, 11971–11976 (1989).

Helical period of Z-DNA

SIR—Pui Shing Ho of Oregon State University, and Michael Youngquist and Robert Driscoll of the California Institute of Technology, have independently pointed out that, in our paper "Scanning Tunnelling Microscopy of Z-DNA" (*Nature* **339**, 484–486; 1989), the apparent helical period in our Figs 1 and 2 is considerably longer than the 43 Å generally accepted for Z-DNA.

Both sets of authors have suggested that the 70 × 46 Å boxed area in Fig. 2a of our paper (enlarged in Fig. 2b) may actually represent nearly two full turns of the helix. In that case, the prominent grooves would probably both be minor grooves corresponding to successive turns of the helix, since the bases in the major groove form a convex surface. This agrees with the observation by Ho that Fig. 1b "shows ten peaks and nine valleys between 0 and 400 Å along the length of the fibre and is consistent with a 41.5-Å spacing of each

vertical displacement." Youngquist and Driscoll point out that "the molecular electrostatic potential surface has a major groove three times wider than the minor groove; the dark bands across the molecule in the STM image, however, are of nearly identical width", as would be the case for the van der Waals surface. They note that the 4–6-Å width of the dark bands accords well with the 3–5-Å width obtained crystallographically. They also point out, correctly, that the Fourier transform frequencies at 20.4, 14.5 and 7 Å in Fig. 1c cannot necessarily be attributed to helical structural features, as they are close to harmonics of the 42-Å period. Such harmonics would be expected since the helix is not a pure sinusoid.

We admit that these suggestions could well be the proper interpretation of our images. We did not notice the discrepancy in dimensions, and were seduced by the excellent visual match between the STM image and the electrostatic model (Fig. 2c). The van der Waals surface, while it gives reasonable correspondence with the grooves in Fig. 2b, does not show the variations in width perpendicular to the helix axis that are evident in Fig. 2b, c. On the other hand, if Fig. 2b is scaled from 70

to 46 Å and projected on Fig. 2c, the fit is remarkably good.

It seems likely that an STM image of DNA could reflect a complex mix of van der Waals and electronic properties. The interpretation of STM images of macromolecules such as DNA involves other issues in addition to the topographical contour, such as the size and structure of the probe tip, its response as it scans along graphite and approaches a thick molecule from the side, its pressure on the sample, and the inhomogeneous electronic structure of the polymer. The most dependable information lies in the periodicities, those features which repeat regularly, regardless of the mechanism of image formation. Judged by this criterion, our images clearly show the left-handed helix and helix-repeat characteristics of Z-DNA.

PATRICIA G. ARSCOTT

GIL LEE

VICTOR A. BLOOMFIELD

D. FENNELL EVANS

Department of Biochemistry,
University of Minnesota,
St Paul, Minnesota 55108, USA
and

Center for Interfacial Engineering,
University of Minnesota,
Minneapolis, Minnesota 55455, USA

Nepotism in the honey bee

SIR—Reports that honey-bee (*Apis mellifera*) workers discriminate nepotistically among nestmate patriline, by Page et al.¹ and others (for example, refs 2–8), have generated considerable interest. But Queller et al.⁹ recently found that *Polistes annularis* wasps, under natural conditions, do not preferentially join more closely related natal nestmates during foundation of their spring colonies. Similarly, within colonies of the polygynous carpenter ant *Camponotus planatus*, workers fail to distinguish their mother queen, sister workers or sister virgin queens from those of other naturally cohabiting nestmate matriline (N.F.C. and S. Cover, unpublished results). The discrepancy between the results obtained in wasps and ants and in the honey bee could reflect a unique kin-recognition ability of the latter. Alternatively, we suggest that the composition of experimental colonies containing artificially low numbers of patriline, of artificially high phenotypic distinctiveness, may yield nepotism as an artefact¹⁰.

Apparent kin recognition can easily arise as the accidental by-product of different discrimination mechanisms based on genetically correlated cues (such as species or mate recognition)¹¹. Social insects are acutely sensitive to foreign odours normally used to identify and reject non-nestmates from the colony. Honey-bee workers in colonies of artificial phenotypic heterogeneity may per-

ceive their peculiar nestmates as smelling more like foreigners than like ordinary nestmates. Discrimination in such behaviours as worker interactions, swarming and queen brood rearing may result, not from a preference for one's own patriline, but from a bias against the other, though not reaching the threshold necessary for rejection^{7,10}. Strange-smelling larvae may in addition elicit discrimination that is normally directed against diseased brood, though again below the threshold for removal; such behaviour is known to be important for disease resistance in honey bees¹².

Various experimental methods have been used to form genetically mixed colonies of honey bees, all of which combine a diversity of phenotypes which may not typically coexist. In early studies, colonies containing multiple kin groups were formed by adopting brood between foreign nests, of known or assumed relatedness^{2,3}. An apparent improvement was introduced by artificially inseminating a queen with drone sperm bearing heritable colour markers, yielding half-siblings visually distinguishable to the investigators (for example, refs 4–8). Because bees recognize relatives by olfaction, the colours themselves were usually assumed not to affect recognition responses. But marker genotypes have been shown to alter the bees' odour phenotypes as well. When a queen heterozygous for a recessive marker is inseminated with sperm

from a single recessive drone, the resulting two worker groups exhibit some preference for nestmates of their own colour, though all are full sisters¹³.

Use of allozyme markers¹ rather than colours does not eliminate the problem without evidence that the variation between genotypes at other loci is comparable to that among offspring of a naturally mated queen. Particularly if the marker-carrying lines used were originally drawn from different populations, they may well differ in a variety of metabolic traits that exaggerate their olfactory discriminability. In addition, the number of patriline in these experiments has been limited to two or three, the number of genetic markers available. In nature, honey-bee queens reportedly mate with 7–17 males¹⁴, presenting workers with a more daunting discriminative task. One study indicates that the nepotistic behaviour observed in two-patriline colonies disappears in colonies containing seven or eight lines⁸.

The degree of discrimination observed within honey-bee colonies is generally weak, never reaching a 2:1 preference for full sisters. Yet workers that are able to distinguish among patriline should maximize their inclusive fitness by aiding full sisters exclusively, particularly when rearing queen larvae — assuming genotypic equivalence in reproductive value. (Feeble full-sister larvae should not be exclusively preferred to vigorous half-sisters. But variation in brood vigour would never lead multiple worker lines to prefer full sisters; rather, each line should prefer the same larvae.) Selection for the collective efficiency of all lineages has been suggested as a brake on kin-group selfishness within the colony¹, but it is unnecessary to invoke adaptation on the colony level to account for the pattern. Biases are expected to be weak when kin recognition is not itself under selection, but is a by-product of another genetic discrimination system¹¹. By contrast, stronger patrilineal differences are reported in studies of genetic influence on honey-bee division of labour, a phenomenon which has also been documented in colonies containing naturally mated queens¹⁵.

Finally, kin recognition can also be induced within ant colonies of artificial phenotypic heterogeneity. When *Camponotus floridanus* workers originating from distant locales (Tallahassee and the Florida Keys) are experimentally adopted into one nest, they detect and persistently investigate nestmates from the other population, with a bias of 1.12:1 (ref. 16). Consistent with the hypothesis that discrimination represents not nepotism, but a side-effect of inter-colony hostility, workers use their recognition ability to bias weak aggression toward alien nestmates, but fail to prefer sisters in coopera-

tive food exchanges and grooming. Given the apparent absence of kin recognition in natural colonies of ants and wasps, and the inconclusiveness of the evidence for honey bees, we would argue that adaptive nepotism among nestmates has still not been demonstrated in any social insect.

NORMAN F. CARLIN
PETER C. FRUMHOFF

*Museum of Comparative Zoology
Laboratories,
Harvard University,
Cambridge, Massachusetts 02138, USA*

PAGE *ET AL.* REPLY—Carlin and Frumhoff suggest that a reduced number of subfamilies ('patriline' derived from different fathers) may lead to atypical behaviour, citing the study of Hogenboom and Velthuis⁸ as evidence. Hogenboom and Velthuis established colonies that had two distinguishable subfamilies, yellow and black, and other colonies that had seven or eight subfamilies, one yellow and six or seven black. They then observed the feeding and 'aggressive biting' interactions of individuals within these colonies to determine if food was being preferentially exchanged among members of the same subfamily and if aggressive biting occurred between individuals of different subfamilies.

We calculated biases directly from the marginal totals of their contingency tables; in all four trials of two subfamilies and three of four trials of seven and eight subfamily colonies, the biases were in the direction expected if preferential kin discrimination was occurring. The bias was statistically significant for one trial ($P < 0.01$) of an eight-subfamily colony while the other three trials lacked sufficient statistical power to test any effect because sample sizes were too small (8, 22 and 3 observations of yellow–yellow worker interactions, respectively). We believe that Carlin and Frumhoff's interpretation of ref. 8 to show a lack of discrimination in worker interactions is incorrect.

The suggestion that the use of genetic markers affects recognition directly or by linkage with genes that affect recognition, is important and plausible. But Frumhoff in his unpublished study¹¹ did not resolve whether the single, recessive gene marker used, or genes linked to that marker, affected recognition. This kind of linkage effect does not influence the reported kin-recognition studies because the markers and their associated linkage groups are distributed randomly within each subfamily; subfamily composition is determined by the paternal genomes. Linkage is not an issue for subfamily recognition because all paternal genes are linked, a consequence of haplodiploidy.

The real issue associated with genetic markers and kin-recognition studies is whether the markers themselves can affect recognition. Colour and allozyme

markers occur naturally and are polymorphic in most populations. Perhaps the most significant argument against Carlin and Frumhoff is that studies using allozymes, single-gene recessive mutations and polygenic integument colour yield the same kind of result as studies that use naturally mated queens and no markers at all¹⁸.

Genetic manipulation may result in greater genotypic diversity in colonies than would occur under natural conditions, though there is no evidence for this in honey bees. Greater phenotypic diversity may increase the experimental resolution of phenomena that occur at very low levels in colonies with less diversity, rather than introduce artefacts. These low-level effects that have been so frequently reported are themselves intriguing evolutionary puzzles.

Experimental evidence does exist to support the conclusion that worker honey bees can recognize and discriminate among nestmates from different subfamilies under experimental conditions. Although everyone acknowledges the limitations of their experiments, there is no evidence that these limitations introduce experimental artefacts.

ROBERT E. PAGE JR

*Department of Entomology,
University of California,
Davis, California 95616, USA*

MICHAEL D. BREED

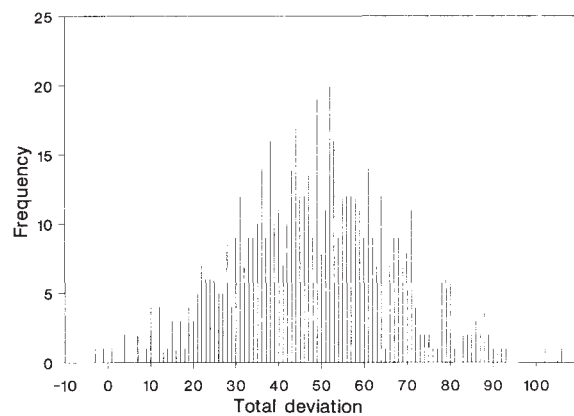
*Department of Environmental, Population
and Organismic Biology,
University of Colorado,
Boulder, Colorado 80309, USA*

WAYNE M. GETZ

*Department of Entomology,
University of California,
Berkeley, California 94720, USA*

SIR—Page *et al.*¹ recently claimed that honey bees working on queen cells bias the rearing of queens in favour of supersisters rather than half-sisters. To test the statistical significance of the finding, Page *et al.* identified "nepotistic" subfamilies (those with highest ratio of adults on queen cells relative to larvae), and summed deviations between expected queen frequency (based on larval frequency) and observed queen frequency of these subfamilies for 30 trials. Page *et al.* then used a computer simulation to generate a population with subfamily frequencies equal to estimates from pooled experimental larvae and queen frequencies, drawing experiment-sized samples for larvae, queens and adults. The simulated nepotistic subfamily was determined as described, and deviation in frequency of these subfamilies between queens and larvae were calculated for 30 simulated trials and summed. The probability of obtaining simulated total deviations as extreme as that observed was estimated as the per cent of 1,000 simulations in which

Simulation was used to generate 600 data sets each of 3 trials of 10 colonies of 3 subfamilies at equal frequency. The frequency distribution of simulations which produced each deviation in 'observed' and expected numbers of queens for nepotistic subfamilies is plotted. Page *et al.* observed a deviation of 60 queens for their experimental data. The figure shows that this is well within the range expected by chance. Data sets were also generated for colonies with differing subfamily relative frequency. Results obtained were similar to those in figure.



the total deviation exceeded 60.

The simulation performed by Page *et al.* supports the nepotism hypothesis, but this probably stems from the interaction of two experimental and two simulation deficiencies. The experimental deficiencies are the small sample sizes and that 6 of the 10 experimental queens were heterozygous for MDH, necessitating allocation of subfamily affinity on the basis of probability. These shortcomings mean that subfamily frequencies were estimated with high sampling error.

The simulation deficiencies are: first, for the real data, the wrong subfamily will be designated nepotistic when, through sampling error, its relative frequency is underestimated in larvae, causing large deviations between observed and expected frequencies in queens. Incorrect subfamilies will be chosen less often in the simulations, because all genotypes are known, and the sample size is larger. Second, worker and queen samples are pooled to provide estimates of simulated subfamily frequencies. (Expected values for the observed data were calculated from larval frequencies alone.) The weighted average is inevitably intermediate between queen and larval frequencies, thereby reducing the magnitude of deviations between frequencies in the simulated queen and larval samples. Thus the total deviation will appear smaller for the simulated data than the observed. If identical procedures are used for the observed data (queen frequencies are predicted from pooled larval and queen frequencies), then the total number of excess queens is -2, indicating no nepotism.

We generated 600 random 'data' sets drawn from a population of three subfamilies at equal frequency and analysed them using the procedure of Page *et al.* (see figure). Of our data sets, 26 per cent had deviations exceeding 60. We then 'estimated the probability of getting a deviation that was as great or greater than our results', using the Page *et al.* simulation. Of our data sets, 84 per cent produced a 'significant' deviation. Thus even data drawn from a totally non-nepotistic population can exhibit 'statistical significance' with these procedures.

There is no evidence in the data of Page *et al.* that subfamily relative frequencies differ significantly between workers and queens, which they should if nepotistic rearing occurred. It is suggested that nepotistic subfamilies are genetically predisposed to rear queens. But nepotistic subfamily identity varies within colonies, suggesting that genetic determination of nepotism does not exist.

BENJAMIN P. OLDROYD
THOMAS E. RINDERER

*Agricultural Research Service,
US Department of Agriculture,
Baton Rouge, Louisiana 70820, USA*

STEVEN M. BUCO

*Statistical Resources,
7332 Highland Road,
Baton Rouge, Louisiana 70808, USA*

PAGE AND ROBINSON REPLY—The ability of worker honey bees to discriminate among queen larvae on the basis of kinship has been controversial since two papers were published in 1984, when R.E.P. and Erickson reported¹⁹ that workers could distinguish between nestmate and non-nestmate larvae while Breed, Velthuis and G.E.R.² failed to find evidence for discrimination. These and other subsequent studies (for example, refs 3, 4) can certainly be criticized on the basis of some artificiality introduced by the experimental methods, criticisms usually stated by the authors themselves (see ref. 17).

Oldroyd *et al.* correctly point out that our demonstration¹ of nepotistic queen rearing is the result of a sampling bias in the Monte-Carlo simulation we used. We have now reanalysed some of our original data using more conventional statistical methods that are free of the biases present in the simulation model. Full details are available on request from R.E.P.

Behavioural heterogeneity within honey-bee colonies is in part a consequence of colony genetic structure. In our study, this was most dramatically demonstrated by the difference in the subfamily composition of the adult workers sampled on queen cells (presumed to be engaged in queen care) and the composition of samples of worker larvae (presumed to be

the pool from which larvae were drawn for workers to select among as queens, or remove). This heterogeneity probably results from patterns of sperm use by queens resulting from the incomplete mixing of sperm from their many mates, or from behavioural biases that are a consequence of the genotypes of workers. The effect of genotype is demonstrated by comparing the subfamily representation of adult workers sampled on queen cells versus those sampled on worker brood. Differences in these distributions are significant in 4 of 10 colonies and highly significant for all colonies combined (our unpublished data, available from R.E.P.).

Worker larval samples differ from queen samples, but these differences are only marginally significant. Both queen and worker larval samples fluctuate significantly from trial to trial within colonies, but these fluctuations are different, suggesting that some additional nonrandom process is affecting the subfamily representation of queens. We suggest that this process involves discrimination during queen rearing.

Our results show that honey-bee colonies have a genotypic and behavioural structure that could favour nepotism during queen rearing, provided that larvae are genotypically labelled, and workers have kinship-related information about larval labels. These necessary elements have all been reported elsewhere¹⁷. But a conclusive demonstration of nepotistic queen rearing remains elusive.

ROBERT E. PAGE JR

*Department of Entomology,
University of California,
Davis, California 95616, USA*

GENE E. ROBINSON

*Department of Entomology,
University of Illinois,
Urbana, Illinois 61801, USA*

1. Page, R.E., Robinson G.E. & Fondrk, M.K. *Nature* **338**, 576–579 (1989).
2. Breed, M.D., Velthuis, H.H.W. & Robinson, G.E. *A. ent. Soc. Am.* **77**, 737–739 (1984).
3. Visscher, P.K. *Behav. Ecol. Sociobiol.* **18**, 453–460 (1986).
4. Noonan, K.C. *Ethology* **73**, 295–306 (1986).
5. Getz, W.M., Bruckner, D. & Parisian, T.R. *Behav. ecol. Sociobiol.* **10**, 265–270 (1982).
6. Evers, C. & Seeley, T. *Anim. Behav.* **34**, 924–925 (1986).
7. Frumhoff, P.C. & Schneider, S. *Anim. Behav.* **35**, 255–262 (1987).
8. Hogendoorn, K. & Velthuis, H.H.W. *Naturwissenschaften* **75**, 412–413 (1988).
9. Queller, D.C., Hughes, C.R. & Strassmann, J.E. *Nature* **344**, 388 (1990).
10. Carlin, N.F. *Neth. J. Zool.* **39**, 86–100 (1989).
11. Grafen, A. *Anim. Behav.* **39**, 42–54 (1990).
12. Rothenbuhler, W.C. *Am. Zool.* **4**, 111–123 (1964).
13. Frumhoff, P.C. thesis (Univ. California, Davis, 1987).
14. Winston, M. *The Biology of the Honey Bee*. (Harvard Univ. Press, Cambridge, 1987).
15. Calderone, N.W., Robinson, G.E. & Page, R.E. *Experientia* **45**, 765–767 (1989).
16. Carlin, N.F., Hölldobler, B. & Gladstein, D.S. *Behav. ecol. Sociobiol.* **20**, 219–227 (1987).
17. Page, R.E. & Breed, M.D. *Trends ecol. Evol.* **2**, 272–275 (1987).
18. Breed, M.D., Butler, L. & Stillerr, T.M. *Proc. natn. Acad. Sci. U.S.A.* **82**, 30058–3061 (1985).
19. Page, R.E. & Erickson, E.H. *A. ent. Soc. Am.* **77**, 578–580 (1984).

Odder than we can think

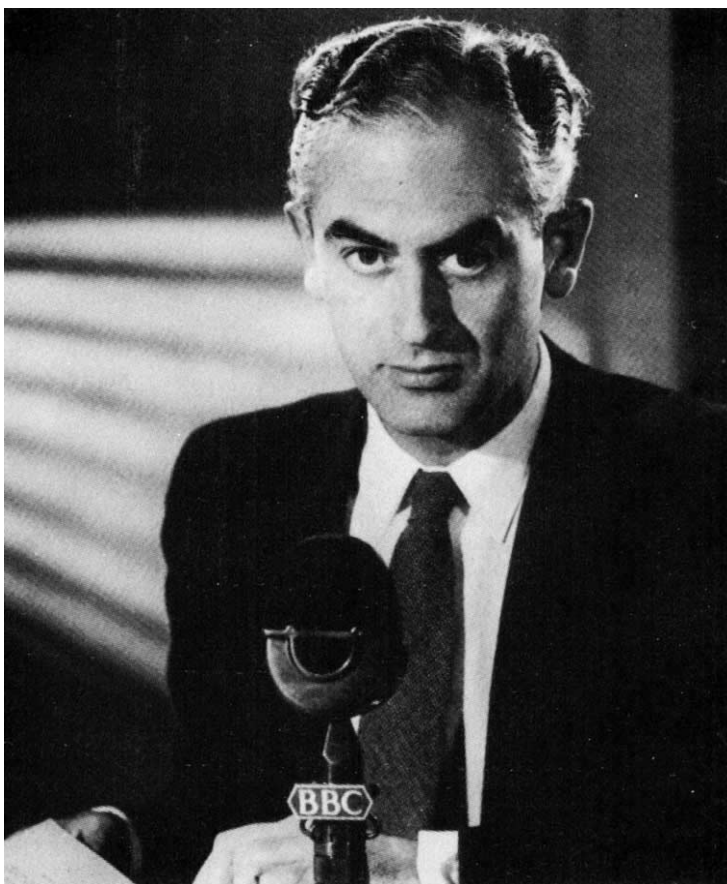
Alex Comfort

The Threat and The Glory. By Peter Medawar. Oxford University Press: 1990. Pp.291. £15. Published in the United States by Harper Collins, \$22.50. **A Very Decided Preference.** By Jean Medawar. Oxford University Press: 1990. Pp.256. £15. Published in the United States by Norton, \$19.95.

ON AN afternoon in the 1960s, Macfarlane Burnet gave a lecture at the British National Institute for Medical Research chaired by the director, Sir Peter Medawar. The lecture was brilliant but discursive — and long. When Medawar rose to thank the speaker he expressed his appreciation, and proceeded to summarize the lecture in five minutes flat. There will be many who had the privilege of knowing and working with Medawar who will recognize in that incident the mixture of precision and quiet irreverence which made him such a notable communicator of science. He once pointed out that compared with artistic creators, scientific discoverers are expendable because their insights are perpetually superseded, and it is for that reason that as his discoveries, which laid the foundations for transplantation surgery, recede into the history of science, his general and philosophical writings come to the fore. *The Threat and the Glory*, published today, together with a new biography, contains both his Reith lectures on the future of man (when the title was proposed to him, he told us he would lecture on the future of man and the inadequacy of academic salaries) and several essays originally written as reviews. Collections of reviews often fail to work, but Medawar never reviewed anything without producing a polished essay which, like Macaulay's reviews, often had a higher durability than the work which occasioned it.

Medawar the researcher was comprehensively recognized in his lifetime, though the concentration on his immunological work sometimes obscures his scope — he was a fundamental theoretician, for instance, on the nature of ageing. Medawar the philosopher of science is now best known for his championship of Karl Popper and of Whewell, and for his summary exposure of the pretence that

science proceeds by induction, which would be "intellectually disabling if it were true". Few eminent biologists of Medawar's time were as philosophically literate: called away from interviewing a raw candidate who wanted to work with



"Isn't he beautiful?" — attracting students to science.

him on the nature of ageing, he would ask the victim to glance at a new book by Bradley and, on resuming the interview, inquire what he thought of it. Medawar's demolition of Teilhard de Chardin, not so much for what he said as for the uncritical way he said it, is a polemical classic.

Herein lies something of a mystery, which none of the published essays solves. As a philosopher, Medawar addresses many of the ethical problems of science — notably the gung-ho baconian view that science exists to do anything which is possible because it is possible, an idiocy which represents the "threat" mentioned in the title. Medawar's position is one of liberal humanism, which represents the "glory". Much of his work was concerned

with science as the delimiter of the imagination in the light of the reality principle: hence his antipathy to mystics such as Teilhard who fail in attempting to square their mysticism with science, and to the "lava flow of ad hoc explanations" which disfigure freudian psychoanalysis. He enjoyed Thomas Reid, the slightly Alf Garnett-like father of "commonsensism", but he also studied Kant.

Yet his final public posture, which differs little superficially from that of T. H. Huxley, leaves a feeling in the reader of a step not taken: never more, perhaps, from a man whose interest in philosophy was kindled by reading Whitehead, than in the fact that nowhere in his writings does Medawar address the philosophical issues now raised in the laboratory by quantum mechanics. Niels Bohr said that anybody, any philosopher of nature in particular, who was not shocked by this subject had failed to understand it. As a serious mathematician, Medawar clearly understood. Probably, with the awful example of Teilhard, an acute awareness of the imaginative human need to become "bunkrapt" and the obligation of limiting this tendency by science, he was still looking for the right words. One of the profound losses inherent in Medawar's death is that one cannot now discuss this area with him and hear how he would interpret its application to the philosophy of science, and to biology in particular. It was left to his prickly and equally rationalist colleague J. B. S. Haldane to suggest that "nature is not only odder than we think, but odder than we can think".

It was said of Reid that his life gave little occasion for biography. Medawar suffered a severe stroke at the age of 54 yet fought back to resume his work (much as did Louis Pasteur), and despite a devastating series of recurrences and the loss of an eye to glaucoma. This is an epic of personal endurance which he wryly summarized as "a very decided preference" for staying alive.

One of Medawar's earliest encounters with Jean Taylor, his future wife, led him to offer to give her tutorials in philosophy. Everyone who knew the Medawars will confirm that her assessment of her own role in the book she has written about their life together, that it was "limited to enabling", understates her contribution. Despite Medawar's intellectual brilliance

From *A Very Decided Preference*.

and his impressive appearance ("isn't he beautiful?" said a visiting lady biologist from Italy) there was a certain diffidence about him and a sensitivity which he kept well concealed from professional colleagues. It was the presence of Jean Medawar and the great understanding between them which provided the steel. Medawar could joke that it was the fortunate possession of an entirely happy marriage which gave him more time than certain of his colleagues to concentrate on work, but the joke was true.

It is hard — and disadvantageous with reviewers — to write a loving biography, particularly of a Nobel prizewinner who lends himself on so many counts to hero worship. But Lady Medawar succeeds brilliantly: she has written an unsentimental tribute to their relationship which is also genuinely revealing of the man — a man who thought that Wagner's plots were "all rot" but was transported by the music, an immensely significant footnote to his grasp of the imaginative as a source

of insight, developed in his philosophical essays. Einstein played the violin, but I have the impression that, like a grounding in philosophy, this Blakean receptiveness is in short supply among biologists.

There cannot be too many extant accounts of a relationship which was happy, passionate and intellectually reciprocal on quite this scale. Moreover, in view of Medawar's "very decided preference", his single-minded determination to press on regardless despite the torpedoes, and his complete lack of any expression of self-pity in the face of massive adversity, it is most fitting that we should have this account of his illness from the woman who shared it and supported him. There have been a few great scientific love stories — the Curies come to mind — but in chronicling this one, Lady Medawar adds to our knowledge of a great man and increases our admiration for a great lady. □

Alex Comfort is at the Windmill House, The Hill, Cranbrook, Kent TN17 3AH, UK.

Chaperones of the cell

R. John Ellis

Stress Proteins in Biology and Medicine. Edited by R. I. Morimoto, A. Tissières and C. Georgopoulos. Cold Spring Harbor Laboratory Press: 1990. Pp. 450. \$97.

ALL cells contain groups of highly conserved proteins that increase rapidly in concentration when the cells are exposed to environmental stresses, including those that cause human disease. The most studied stress is high temperature. There is much indirect evidence which suggests that stress proteins have a protective function, allowing cells both to recover from the inducing stress and to survive subsequent stronger stresses that would otherwise be lethal.

The study of stress (or heat-shock) responses at the molecular level has, until recently, concentrated on the mechanisms that cause these stress proteins to accumulate. A significant component of these mechanisms is the rapid increase in transcription of genes encoding stress proteins. Much detailed information is available about the regulatory DNA sequences and DNA-binding proteins involved, although the mechanism by which stresses are sensed to trigger transcriptional activation is not known. Even more unclear is the function of the diverse groups of stress proteins. The high degree of conservation of the amino-acid sequences among these groups in all organisms, together with the observation that many of the proteins are present when the organism is not subjected to stress, suggest that these proteins

have functions essential to normal cellular operations but are required to a higher degree under stress conditions. In the past few years this area of stress-protein research has been stimulated by increasing evidence suggesting that many stress proteins function as molecular 'chaperones'.

Chaperones are a distinct family of proteins required in certain cellular processes such as protein synthesis and assembly, protein transport across membranes, and the functioning of oligomeric protein complexes such as those involved in DNA replication and the recycling of endocytotic vesicles. All these processes produce changes in the state of protein folding and/or oligomerization, and so all involve the transient exposure of interactive surfaces to the intracellular environment (interactive surfaces are any regions of intra- or intermolecular contact important in maintaining the functional structure). Such exposed surfaces run the risk that they may interact 'incorrectly' with one another to produce non-functional three-dimensional structures. The probability of incorrect interactions occurring is known to vary widely but to be increased at high protein concentrations and high temperatures.

Molecular chaperones function by recognizing and binding to such surfaces to form stable complexes in which incorrect interactions are inhibited. These complexes are then dissociated by other proteins, often using the energy of ATP hydrolysis, under circumstances where incorrect interactions are favoured. Chaperones convey no steric information for either protein folding or protein oligomerization, and neither bind to nor are components of the final functional structures — hence the term chaperone.

This concept can readily accommodate the functions of stress proteins if it is assumed that the primary effect of stress is to cause the appearance of interactive surfaces that are recognized by chaperones. For example, this model suggests that heat shock causes proteins to denature and form incorrect aggregates, whereas heat-shock proteins inhibit these processes by binding to the interactive surfaces produced by high temperature. Thus the stress response amplifies a pre-existing function which all cells require for their operation under non-stress conditions.

The recent evidence that supports this interpretation forms the more interesting part of this collection of 18 chapters covering many aspects of stress proteins in microbial and animal cells, including diseased human cells. As well as discussion of the molecular details of stress proteins, their genes and their mode of action as chaperones, there are chapters devoted to more organismal aspects such as the physiological adjustments to fluctuating thermal environments, the febrile response, immunological consequences of stress and the use of hyperthermia in cancer therapy. The quality of discussion and presentation is high, and each chapter is well-referenced — it would be hard to find a better introduction to the latest thinking in this field.

A final heretical thought — perhaps all stress proteins will turn out to be molecular chaperones (the converse is not true), in which case we should drop the terms 'stress' and 'heat shock', as these words refer to only one specialized aspect of the much broader, essential roles of these proteins. □

R. John Ellis is in the Department of Biological Sciences, University of Warwick, Coventry CV4 7AL, UK.

New in paperback

■ **Nitrogen Fixing Organisms**, by Janet Sprent and Peter Sprent, is a general introduction to the subject for undergraduate and graduate students as well as providing a guide to theory and practice for researchers in agriculture, forestry and environmental management. Publisher is Chapman and Hall, price is £14.95.

■ Freeman has issued a second edition of **Electromagnetism: Principles and Applications** by Paul Lorrain and Dale Corson. The book is an abbreviated version of the more advanced **Electromagnetic Fields and Waves** by the same authors. Price is £24.95.

■ David Suzuki and Peter Knudson's best-seller **Genetics** is available in paperback. The authors derive a set of ethical principles to guide genetic engineers and others through this "explosive" area of research. Publisher is Unwin Hyman, price is £6.99.

■ **Chemical Bonding Theory** by Brian Webster introduces modern ideas about the chemical bond, with emphasis on molecular orbital theory. Publisher is Blackwell Scientific, price £14.95. (Also available in hardback at £29.50). □

The importance of being alpha

W. C. McGrew

The Chimpanzees of the Mahale Mountains. Sexual and Life History Strategies.

Edited by Toshisada Nishida. University of Tokyo Press: 1990. Pp. 328. £60.

IF ASKED about their images of wild chimpanzees, Europeans and North Americans would probably agree on the importance of Jane Goodall's contribution. Her long-term study of the Kakombe community of apes in western Tanzania means that, after 30 years of research, individuals like Flo, Fifi, Goblin and so on are well known to millions of lay people as well as to academics. This is as it should be, for more than any other person she, with the help of the National Geographic Society, has defined our Western view of wild chimpanzee nature.

But another study of wild chimpanzees in western Tanzania is arguably equally important, if not more so. This is the 25-year-long project of Toshisada Nishida and his colleagues at Kasoje in the Mahale Mountains. Their research programme, which is based at Kyoto University, has received wide acclaim in Japan, but is not well-known in the West, except among primatologists. One reason is, that unlike Goodall with *In the Shadow of Man* (1971) and *The Chimpanzees of Gombe* (1986), Nishida had not produced a book in English. That long-awaited convenience has now been supplied.

Differences in structure make direct comparisons between Nishida's volume and Goodall's impossible. While the latter has always written integrated, personal accounts, Nishida has edited an anthology

of 16 chapters by 10 contributors. While Goodall has always striven for comprehensiveness, Nishida *et al.* are here more selective, as the subtitle suggests.

Finally, Goodall's forte has always been a blend of primarily qualitative, and only secondarily quantitative, description that focuses on individual apes. This richness is matched in the present volume only by Itani's chapter on his safari surveys through the Mahale Mountains and their hinterlands during the past 25 years. It is engrossing, adventurous stuff, and reiterates a technique that still has much to offer.

Of the book's other chapters, two stand out. One is Nishida's opening overview of 2 years of research, and is the single most accessible source in English on the history of Japanese primatological efforts in the region. Also exceptional is the compilatory chapter on demography and reproductive profiles. This is where longitudinal study of a long-lived species comes into its own. Impressive (such as a median birth interval of as long as 72 months) and unexpected (such as bimodal annual birth-peaks) results appear on almost every page of this chapter, which is an actuarial goldmine.

I have insufficient space to do justice to the 12 'data chapters', each of which is essentially equivalent to a meaty journal article. Their topics range from the specific (social context of the pant-grunt vocalization, age differences in ant-eating) to the more general (social relations among adult males, mother-infant relationships).

Comparisons between Gombe and Mahale highlight the unique aspects of the latter: at Mahale, the main study group, M-Group, is the biggest ever recorded for the species, with more than 100 members. This has numerous implications — for example, the joint strategy of a male and female of disappearing on a temporary 'honeymoon safari', as first detailed by

Caroline Tutin at Gombe, cannot occur at Mahale. There are just too many male rivals to be evaded. Consequently the possessive tactics of the alpha male assume disproportionately greater importance. He dominates the matings of females at the time of ovulation. Again and again, the importance of being 'the alpha' (highest-ranking male) comes through. The long-time holder of this position, Ntologi, deserves just as much fame in his own right as do Washoe or Lana in theirs.

My only complaint is one of frustration. Nishida has told only a fraction of the story, and whole areas of interest like tool use are largely ignored. Is it too much to ask him to give us a fully synthetic sequel? □

W. C. McGrew is in the Department of Psychology, University of Stirling, Stirling FK9 4LA, UK.

Ordered growth

Lewis Wolpert

Pattern Formation. Ciliate Studies and Models. By Joseph Frankel. Oxford University Press: 1990. Pp. 314. £52, \$65.

CILIATES are bad and good for cell biologists. Bad because they undermine our self-esteem by showing how little we understand about cell structure at the molecular level; good because they offer us problems, which, when solved, will be a quantal leap in our understanding of not only cells but, surprisingly, embryos. For among the extraordinary properties of ciliates is that their mechanism of pattern formation and regulation is formally similar to that of multicellular systems.

The varied surface patterns of ciliates are highly complex and asymmetrical, yet are essentially a variation on the organization of ciliary units. Division of ciliates is a good introduction to the basic problems. When a ciliate like *Tetrahymena* divides, the process is more like budding in hydra than a typical division of a cell. In fact its growth and division is best likened to a clonal cylinder. The ciliate grows longitudinally, and before division a second set of structures develops in the posterior half. In the fission zone, in the middle of the cell, two opposite poles of the cell are thus juxtaposed. The oral apparatus defining the anterior end of the new cell begins with the formation of an anarchic field in which the ciliary units are so oriented that no order is evident. Gradually, the highly ordered organelle emerges. In addition to this local ordering there is long-range order which places the organelles in the correct relationship to each other.

Ciliates provide the best example of structural inheritance — that is, pre-



The plunge of the diver from this world to the next. The platform symbolizes the pillars of Herakles, the edge of the known world, and the water the unknown, limitless ocean beyond. The painting is on the underside of a coffin lid at Paestrum, one of the best preserved classical cities of the ancient world. From *Paestrum* by John Griffiths Pedley. Published by Thames and Hudson as part of the series "New Aspects of Antiquity", general editor Colin Renfrew. Price is £20, \$35. □

existing structural organization can be inherited during their vegetative growth and division, and is not under genetic control. This is one of the very few examples of inheritance not involving an alteration in DNA. The classic example is the inheritance of ciliary rows that have been inverted by 180°. But it is not just inversion of ciliary rows that can be inherited. Large-scale reversals of asymmetry are possible so that there is a change in intracellular handedness that affects the arrangement of all the cortical structures within its domain.

Ciliates show a remarkable capacity for regeneration and pattern regulation. It is as if the cell is behaving like a classical embryonic gradient field. Frankel has proposed a formal model, based on positional information, which attempts to account for a large number of results. Treating the ciliate as a cylinder, it is assumed that there are both longitudinal and circumferential positional values giving, in effect, a two-dimensional coordinate system, which makes the different parts of the cortical system non-equivalent. One property of the system is its tendency to achieve a continuum of normally spaced positional values, and in this it is similar to the polar coordinate model that has been so successful in accounting for the regeneration of multicellular systems. A key feature of the model is that positional values are 'interpreted' by the development of specific structures. Both positional value and interpretation are, however, completely without a molecular basis. But as Frankel makes clear, the model provides a partial explanation for a large amount of information which would otherwise be incomprehensible. Moreover, there are few errors of commission, that is, a structure forming where it is not expected.

This brief review does little justice to the richness of the material, the clarity of the exposition, or how well Frankel relates ciliate patterning to that in multicellular organisms. He has written an outstanding book. One can only hope it will encourage others to enter the field. The difficulties in solving the problem are great, but so are the rewards. One cannot but believe that underlying patterning in the single cell is a fundamental biological mechanism that has been taken over by multicellular organisms. □

Lewis Wolpert is Professor of Anatomy, University College and Middlesex School of Medicine, Cleveland Street, London W1P 6DP, UK.

■ Two new books from Cambridge University Press also address early development. *This Side Up* by Robert Wall, describes spatial determination in the early development of animals. Price £70, \$110. *Morphogenesis* by Jonathan Bard, is about the cellular and molecular processes of developmental anatomy. Price £35, \$54.50. □

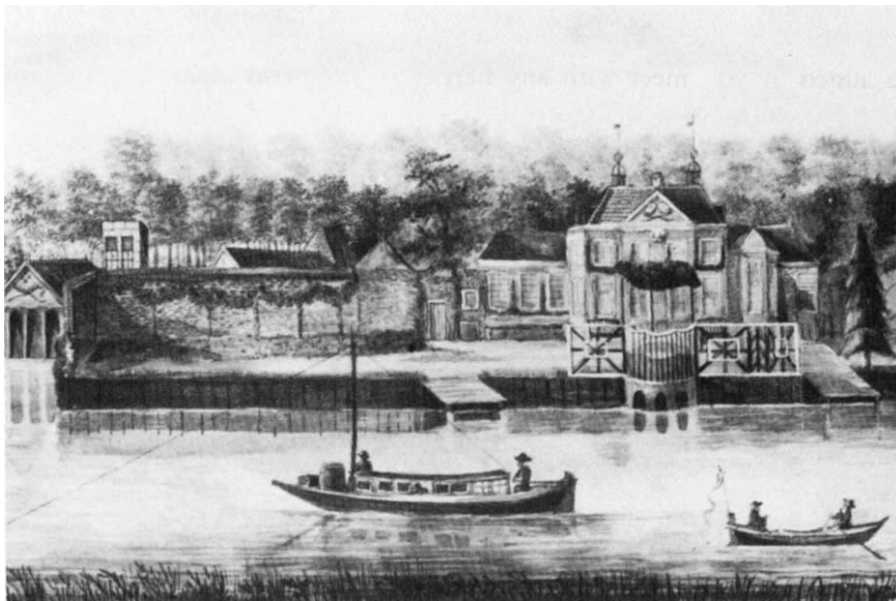
Early blooming

Ghilleen T. Prance

The Chelsea Gardener. Philip Miller 1691–1771. By Hazel le Rougetel. *Natural History Museum Publications, London: 1990. Pp. 212. £14.95, \$29.95.*

"MILLER'S *Gardeners' Dictionary* is the most important horticultural work of the eighteenth century" says Professor William T. Stearn, who contributed the final chapter to this book. This is certainly true; indeed, Philip Miller was also the leading British gardener of his century. As the keeper of the Apothecaries' Physic Garden at Chelsea from 1722 to 1770, he built up the most richly stocked garden in Europe. Miller was both botanist and horticulturalist and his dictionary is the

his correspondents were Sir Joseph Banks and Sir Hans Sloane, and a fascinating chapter deals with his transatlantic correspondence, particularly with plant collector John Bartram. Another chapter is about Miller's most valued flowering shrub, the rose. The Provence appears to have been Miller's favourite among the many species and varieties of roses that he cultivated at Chelsea. Miller did not confine his horticulture to the garden directly under his care in Chelsea but advised and exchanged material with many of the great estates of England, such as the garden of the third Duke of Argyll at Whitton which accumulated a large collection of trees. On the death of the duke in 1761, many of the beautiful trees at Whitton were transferred by Lord Bute to Princess Augusta's garden at Kew, the precursor of the Royal Botanic Gardens. Fortunately, Miller's work at Chelsea has been continued by sub-



Meerburg, the house and garden of M. de la Court. Philip Miller visited this garden, at Oud-Poelgeest, Holland, in 1727, and was instructed by its gardener on the cultivation of pineapples. (From *The Chelsea Gardener*.) □

source of the original descriptions and names of many species of plants.

Miller was a contemporary of Linnaeus and, in the eighth and last edition of his great work, he adopted the Linnaean system of nomenclature which he had earlier resisted. He was a great correspondent and maintained contact with botanists and plant collectors from all around the world. As a result, he obtained for the Physic Garden many species from Europe, North America, South Africa and the West Indies which were not seen by Linnaeus. Miller named these both from living material and from herbarium specimens, so his dictionary is a work of major importance to taxonomic botany. Many familiar garden plants still bear the name of Miller after their Latin binomials.

This delightful new book about Miller covers many aspects of his life and his influence on the world of botany. Among

sequent generations and the Physic Garden is still a testimony to its greatest keeper. This book will make any visitor's trip to today's garden even more interesting.

The book is abundantly illustrated with 16 colour plates and numerous black-and-white drawings carefully selected from Miller's *Gardeners' Dictionary* and *Figures of Plants* and engravings of Chelsea. It is attractive and well laid out, and will be of interest both to botanists and to horticulturalists who are interested in the history of their discipline. □

Ghilleen T. Prance is Director of the Royal Botanic Gardens, Kew, Richmond, Surrey TW9 3AB, UK.

■ The Chelsea Physic Garden, 66 Royal Hospital Road, London SW3 4HB, UK, is open to the public on Wednesdays and Sundays from 2 p.m. to 5 p.m. □

Sun and dust versus greenhouse gases: an assessment of their relative roles in global climate change

James E. Hansen & Andrew A. Lacis

Many mechanisms, including variations in solar radiation and atmospheric aerosol concentrations, compete with anthropogenic greenhouse gases as causes of global climate change. Comparisons of available data show that solar variability will not counteract greenhouse warming and that future observations will need to be made to quantify the role of tropospheric aerosols, for example.

OBSERVATIONS of steadily increasing concentrations, principally man-made, of greenhouse gases in the Earth's atmosphere¹⁻³ have led to the expectation of global warming during coming decades⁴⁻⁶. Computer simulations, supported by palaeoclimate studies, suggest that the potential greenhouse climate change within a century could rival the difference between today's climate and the last great ice age of 20,000 years ago^{7,8}. But the greenhouse effect is in competition with other mechanisms for climate change. Solar variability, although a speculative subject, has received much attention, and recent observational advances allow initial quantitative comparison of the effects of the Sun and the greenhouse. Another mechanism, a change in the atmospheric aerosols (natural and man-made), is of comparable importance.

Because curtailing the growth of greenhouse gases would require significant changes in global energy use and unprecedented international cooperation, it is essential to develop a good quantitative understanding of the relative importance of these different mechanisms of climate change. We also need to know how the magnitudes of such forced climate changes compare with unforced internal fluctuations of the climate system. Here we compare the above climate forcings, describe observations that would help sort out cause and effect of near-term global climate trends, and discuss implications for policy making.

Global climate can fluctuate without any change in the external forcing. For example, the Earth's simulated mean surface temperature varies as much as 0.4 °C in a 100-year run of a global climate model⁹ with fixed solar irradiance and fixed greenhouse gases (Fig. 1). These fluctuations arise because the coupled nonlinear equations describing atmospheric structure and motion have solutions exhibiting chaotic behaviour. Tiny perturbations or changes of initial conditions (the flap of a butterfly's wings) give rise to solutions that are different on the timescale of months or years, and these fluctuations are magnified on decadal timescales as a result of the thermal inertia of the ocean, even if possible changes of ocean circulation are ignored^{10,11}. In effect, the atmosphere and ocean do a lot of 'sloshing' around. Some of the sloshing, for example, the El Niño/Southern Oscillation phenomena¹², will eventually be predictable on a limited timescale, but most of it is of a chaotic nature for which long-term prediction is only possible in a statistical sense. An externally forced climate change must be at least comparable in magnitude to this internal noise to be clearly discernable.

A climate forcing, natural or anthropogenic, is an imposed change that modifies the planetary radiation balance, thus affecting the planetary temperature. The natural forcings that seem to be most significant, based on systematic comparison of radiative effects, are changes of stratospheric aerosols owing to large volcanic eruptions and changes of solar irradiance. The largest

anthropogenic forcings seem to be increasing infrared-absorbing (greenhouse) gases, man-made tropospheric aerosols, and perhaps changes of surface reflectivity owing to desertification and deforestation.

The climate change that results from a change in the climate forcing depends on many feedback processes in the climate system. These feedbacks, including changes of clouds, water vapour, ice and snow cover, are complex and not well understood. Thus, climate sensitivity is very uncertain; for example, it is estimated that doubling the concentration of CO₂ in the atmosphere could lead to an eventual global warming anywhere in the range 1.5–5.5 °C^{4,8,13,14}. Fortunately, this uncertainty can be largely avoided by contrasting the magnitude and timescale of the climate forcings rather than the resulting climate response. High-frequency changes in the forcing have less impact than a sustained forcing, because of the thermal inertia of the climate system. But uncertainties about ocean mixing rates, and thus about the effective thermal inertia of the system, only provide further reasons to focus first on climate forcing, and then discuss climate response.

Comparison of solar and greenhouse effects

A change of solar irradiance is perhaps the simplest climate forcing, and it serves as a standard for comparison. At the Earth's mean distance from the Sun, the irradiance normal to the Earth–Sun line is $\sim 1,370 \text{ W m}^{-2}$. Because the Earth's surface area is four times its cross-section and 30% of the sunlight is reflected back to space without being absorbed, the mean solar heating of the Earth is $\sim 240 \text{ W m}^{-2}$. A change of solar irradiance by 0.1% is therefore a climate forcing of $\sim 0.24 \text{ W m}^{-2}$.

Greenhouse and solar climate forcings of recent years can be compared accurately. Precise measurements of CO₂, the

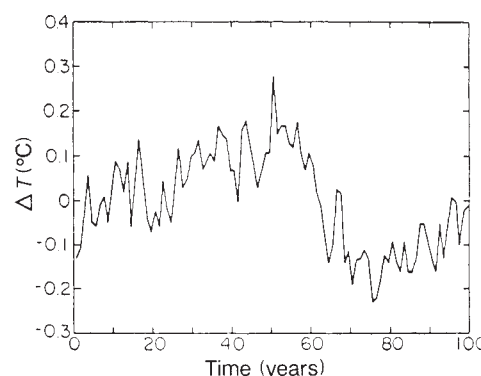


FIG. 1 Global temperature in 100-year run of a climate model⁹ with no variations of climate forcing.

principal anthropogenic greenhouse gas, were initiated in 1958¹. Since then, CO₂ has increased from 315 p.p.m. to over 350 p.p.m. The history of chlorofluorocarbons (CFCs) is also known¹⁵, because CFCs are entirely man-made and production records are available. The trends of methane (CH₄) and nitrous oxide (N₂O), which contribute much less than CO₂ and CFCs to the greenhouse forcing over the past 30 years¹⁵, are known approximately¹⁶. Ozone (O₃), another greenhouse gas, is believed to be decreasing in the stratosphere and increasing in much of the troposphere, but the changes are so variable and poorly measured that it is impossible to say whether they cause a net global warming or cooling^{15,17}. It is unlikely, however, that the ozone global climate forcing is >10–20% of the net forcing by the other gases.

The net climate forcing by CO₂, CFCs, CH₄ and N₂O for the period 1958–1989 is >1 W m⁻² (Fig. 2). This is the heating of the Earth's troposphere calculated with a simple (one-dimensional radiative-convective⁸) or a more sophisticated (three-dimensional⁹) climate model for the indicated changes of these gases. We calculated an increased greenhouse forcing between 1850 and 1989 (Fig. 2) of ~2 W m⁻². Others^{16,18} have reported anthropogenic greenhouse forcing as large as 2.5 W m⁻², the difference being due mainly to small increases of CO₂ and CH₄ before 1850, stratospheric H₂O (which chemical models suggest may have increased because of added CH₄) and uncertain O₃ changes. We have omitted any forcing owing to O₃ or stratospheric H₂O changes because even the sign of the O₃ forcing is uncertain and there are no confirming observations of long-term stratospheric H₂O changes. The 2–2.5 W m⁻² greenhouse forcing is uncertain by a further 10–20% owing to imprecisions in the radiative parameters^{6,19}. Our calculated history for the greenhouse climate forcing since 1958 is shown as the solid curve in Fig. 3.

Precise measurements of solar irradiance, obtained above the interfering effects of the Earth's atmosphere, were initiated in the late 1970s. A cavity radiometer²⁰ on the Nimbus 7 spacecraft has obtained data from late 1978 to the present, and an active-cavity radiometer²¹ on the Solar Maximum satellite with more sophisticated calibrations obtained data from late 1979 through mid-1989. These instruments are believed to be capable of precisions (relative accuracies) of better than 0.1%. Despite some differences in results of the two instruments, they are consistent in showing a decline of solar irradiance of ~0.1% between 1979 and the mid 1980s, with a partial recovery of the irradiance by 1989.

Solar forcing of the climate system based on the present official Nimbus 7 data²⁰ is shown in Fig. 3, with zero forcing defined as the mean for the period of measurement. The combined greenhouse and solar forcing is also shown. Over the common period of accurate solar and greenhouse data the changing Sun

significantly modulated the net climate forcing, but it did not alter the trend.

We stress that climate response to a given forcing depends on the history of that forcing over a period comparable to the response time of the climate system^{8,10}. The response time is of the order of decades, and may be a century or more if climate sensitivity is high^{8,22–24}. Because of this damping, the climate impact of a fluctuating forcing, such as the solar change in Fig. 3, is much less than that of a steady forcing. But solar irradiance may have long-term, as well as short-term, variations.

We also note that the climate system would not be expected to respond in exactly the same way to a change of solar forcing as it would to a change of greenhouse forcing of the same magnitude and timing. Solar heating, for example, is concentrated toward low latitudes, whereas greenhouse heating is more uniformly distributed; also, because solar variability is greatest at ultraviolet wavelengths²⁵, a substantial fraction of the change of absorbed solar energy occurs high in the atmosphere where it is less effective in warming the surface. Tests with global climate models^{8,26} do, however, yield remarkably similar responses for doubled atmospheric CO₂ and for a (spectrally uniform) 2% increase of solar irradiance, both forcings being in the range 4–4.8 W m⁻². Thus a comparison of solar and greenhouse forcings, as in Fig. 3, seems to be meaningful to first order.

The changes of solar irradiance in the past decade are associated with changes in the area of sunspots, which are dark cool regions of reduced irradiance, and faculae, which are bright regions of increased irradiance^{27,28}. The faculae effect dominates, and thus the irradiance increases in periods of maximum solar activity. If the empirical relationships between the areas of sunspots and faculae and the solar irradiance deduced from the data of past decade were valid on all timescales, longer-term changes of solar irradiance would hardly exceed 0.1%. Nor are large fluctuations of energy flow from the solar interior expected: even if the rate of nuclear energy production varied, the long diffusion time for a photon to reach the solar surface, ~10⁴ years, would smooth out the variations²⁹. But changes in the efficiency of convection in the outer convective layers of the Sun, caused, for example, by fluctuations in magnetic field strengths, can alter the rates of energy storage and release. Given the magnitude of surface-layer energy reservoirs, this implies that there may be important changes in the solar irradiance on timescales from decades to centuries²⁹.

Indeed, Eddy³⁰ has argued that reduced solar irradiance during the Maunder minimum of solar activity (1640–1720) may have been responsible for the Little Ice Age when global temperature is estimated to have been as much as 1 °C colder than today³¹. Other possible causes for that climate change exist, including fluctuations of ocean heat transport and the increase

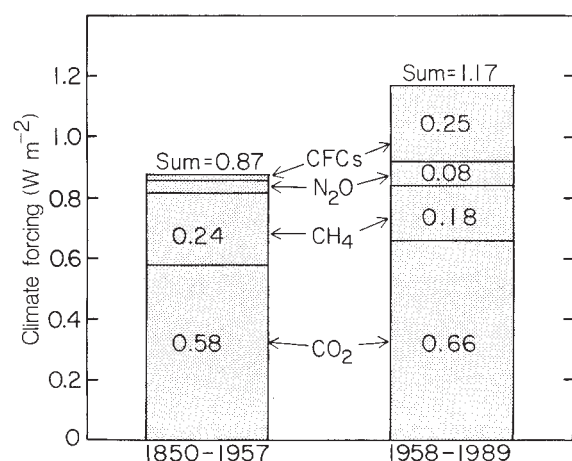


FIG. 2 Added greenhouse climate forcings (W m⁻²) for the periods 1850–1957 and 1958–1989. Assumed gas abundances for the three times (1850, 1958, 1989) dividing the two intervals are: CO₂ (285, 315, 351 p.p.m.), CH₄ (0.8, 1.29, 1.71 p.p.m.), and N₂O (280, 289, 309 p.p.b.). Abundances for the many CFCs and references for data sources are given in ref. 15. Calculated N₂O and CFC forcings for the interval 1850–1957 are ~0.035 and 0.015 W m⁻², respectively.

of greenhouse gases between the 1700s and 1800s, but solar variability is one of the plausible mechanisms. Is it likely that a future decrease in solar activity may cancel greenhouse warming? That tack is taken in a recent report³² to the chief of staff of US President Bush, which contends that solar irradiance can be expected to decline early in the twenty-first century. Thus the authors argued against measures to slow down the increase in greenhouse forcing because such efforts "could turn out to be unnecessary or even harmful if a substantial natural cooling occurs in the twenty-first century".

In comparing possible solar and greenhouse climate changes, one must take care that consistent assumptions are made about climate sensitivity. This can be done most simply by comparing the climate forcings, a simple comparison being valid because the timescales of supposed Little Ice Age forcing and recent greenhouse forcing are similar. Anthropogenic greenhouse forcing has already reached $2\text{--}2.5\text{ W m}^{-2}$, equivalent to an increase of solar irradiance by $\sim 1\%$. Given the increase of greenhouse forcing in the past three decades (Fig. 2), it is apparent that it will reach a level equivalent to a solar change of at least 2% by the middle of next century, unless the rate of growth of greenhouse-gas emissions is reduced.

The possibility of a decline of solar irradiance by 2%, although 20 times larger than the changes measured in the past 11 years, has not been ruled out categorically. The irradiance of some solar-type stars has been observed to change by several tenths of a per cent³³. The Smithsonian Observatory monitored the Sun from mountain tops for the period 1902–1962, initially reporting variations of $> 1\%$. But comprehensive reviews of these data³⁴ concluded that the variations were due primarily to fluctuations of atmospheric transparency, and that they established only an upper limit of $\sim 0.3\%$ for solar variations. Evidence of solar change over a longer period is provided by measurements of the solar diameter. The diameter of the Sun is affected by thermal energy storage in its outer layers because it must maintain global hydrostatic equilibrium on timescales greater than an hour. The small changes of solar diameter measured over the past 275 years are near the limit of detectability³⁵. Although the exact relation between the solar diameter and luminosity is uncertain, there probably were secular luminosity variations during that period, but they did not exceed several tenths of a per cent^{31,36,37}.

These results do not imply that solar variability has been unimportant in past climate change or that the Sun's effects will be negligible in the future. For example, it has been shown that an increase of solar irradiance by only $\sim 0.3\text{--}0.4\%$, just within the range of possibility, is one conceivable explanation of the observed global warmth of the 1930 and 1940s⁵. Also, Wigley and Kelly³¹ have shown that solar variations of several tenths of a per cent may explain many of the climate variations of the past 10,000 years.

Even if there are changes in the solar irradiance in coming decades, it is far from certain whether it will increase or decrease. One commonly used indirect measure of solar activity covering thousands of years is the record of the carbon isotope ^{14}C preserved in tree rings. ^{14}C is continuously formed in the atmosphere by incoming galactic cosmic rays, which are modulated by solar activity. Spectral analyses of the ^{14}C data, when extended into the future, suggest that solar activity will be increasing during the next several decades^{38,39}. Inferences from such proxy records are, however, very uncertain. It is therefore important to compare the irradiance during the upcoming 11-year solar cycle to that of the previous cycle. This will provide the first reliable evidence of the extent to which solar change may add to, or subtract from, greenhouse forcing during coming decades.

We can conclude only that solar irradiance would need to decline by $\sim 2\%$ to counter greenhouse climate forcing by all the anthropogenic gases that will have accumulated in the atmosphere by the middle of next century, assuming no reductions in greenhouse emissions. Although such a solar decline is not

impossible, it is much larger than existing indications of solar variability. Furthermore, solar fluctuations can be positive as well as negative. Measurements of solar irradiance during the next decade or two are crucial for detection of any long-term trends and to help sort out cause and effect of observed climate changes.

Aerosols

Other than greenhouse gases, the largest known global climate forcing is that due to changing atmospheric aerosols. We first discuss stratospheric aerosols, which have a high degree of natural variability and have been well measured, and then tropospheric aerosols.

Stratospheric aerosols arise mainly from volcanic eruptions. Benjamin Franklin⁴⁰ noted that volcanic aerosols reflect sunlight to space, and therefore argued that the eruption of a large volcano on Iceland may have been responsible for unusual cold in 1783–4. Many subsequent studies have found a tendency for eruptions producing a large amount of aerosols over much of the Earth, such as Tambora (1815), Krakatoa (1883) and Agung (1963), to be associated with global cooling of a few tenths of a degree Celsius for a year or two after the eruption. But, because this is only comparable to unforced global temperature fluctuations (Fig. 1), cooling cannot be identified for every large volcano.

In situ stratospheric measurements during the past few decades show that the dominant volcanic aerosols are small droplets of sulphuric acid, which form from volcanic sulphur dioxide and persist a year or more after the eruption. Assuming that the aerosols of earlier volcanoes were also sulphuric acid, estimates of atmospheric transparency can be used to calculate the volcanic aerosol climate forcing^{41,42,43}. The uncertainty in the resulting decadal-mean aerosol forcing is probably less than a factor of two.

The calculated volcanic climate forcing (Fig. 4) at times rivals or exceeds the greenhouse forcing, but the latter clearly dominates the long-term trend. Moreover, volcanic forcing is irregular. As discussed above, brief forcings have much less of an impact than those maintained for several decades. Thus a clustering of several volcanoes is required to have a significant impact on long-term climate. Nevertheless, it is obvious that stratospheric aerosols must be monitored to determine cause and effect of climate change during the next few decades.

Tropospheric aerosols are also an important climate forcing. In the 1970s it was speculated that increasing concentrations of anthropogenic aerosols might send the Earth into an ice age^{45,46}. This speculation was probably fuelled by the fact that the Northern Hemisphere had cooled between 1940 and 1970, as well as by anecdotal information on increases of aerosols in many places. Industrialization, urban pollution, mechanized agriculture, population pressure in semiarid lands—the 'human volcano'—added noticeably to tropospheric aerosol loading. Although in certain circumstances, such as the absorbing haze in the Arctic^{47,48}, anthropogenic aerosols can have a warming effect, the overall direct radiative impact of man-made aerosols is clearly one of cooling⁴⁹.

In the past decade, since it was realized that global temperature was rising^{5,50,51}, tropospheric aerosols have received less attention than the greenhouse effect, but they have not been disproved as an important agent of climate change. Tropospheric aerosols have a global average optical depth $\tau \approx 0.1$ (ref. 52), where τ is defined by the vertical transmission of sunlight, $T = e^{-\tau}$. A 10% increase of τ yields a climate forcing of $0.2\text{--}0.3\text{ W m}^{-2}$, (refs 49, 53), and the effect is nearly linear for feasible changes of τ . Although adequate measurements are unavailable, we estimate that anthropogenic aerosols comprise $\sim 25\%$ of all aerosols, on a global average, implying a climate forcing of $0.5\text{--}0.75\text{ W m}^{-2}$. T. Anderson and R. Charlson (personal communication) argue that as much as 50% of global aerosols may be anthropogenic, implying a forcing of $1\text{--}1.5\text{ W m}^{-2}$. In either

case aerosol forcing is significant, but smaller than the greenhouse forcing of $2\text{--}2.5\text{ W m}^{-2}$. We have argued⁵ that aerosols are not a dominant global climate forcing because there is no evidence for an aerosol trend at remote locations such as Hawaii. Another argument⁵⁴ is that, despite anthropogenic aerosols being concentrated mainly in the Northern Hemisphere, temperature trends are similar in the two hemispheres, at least over the full century^{50,51}. Nevertheless, the direct radiative effect of anthropogenic aerosols must counter greenhouse warming to some degree.

Because of their role as cloud condensation nuclei, a change of aerosols in the troposphere can alter the occurrence and optical properties of clouds. This aerosol effect is complicated, depending on factors such as the supersaturation spectrum of the aerosols⁵⁵, which describes the critical supersaturation at which each aerosol is activated as a condensation centre. It has been found, however, that in most situations added aerosols increase the number of active condensation nuclei and thus increase the cloud particle number density⁵⁶. Twomey⁵⁵ therefore predicted that anthropogenic aerosols would decrease the mean cloud particle size and inhibit rainfall, and that both effects would increase cloud reflectivity.

Although evidence for such phenomena has existed for decades^{57,58}, dramatic reconfirmation of Twomey's expectations was obtained recently from satellite and aircraft measurements^{59–61}, which revealed cases of increased cloud reflectivity in trails behind ships, apparently owing to smokestack aerosols. The phenomenon was identified in shallow stratocumulus clouds under stable meteorological conditions, and was shown to involve decreased particle size⁶⁰ and decreased rainfall⁶¹. Qualitatively, the increased cloud reflectivity must cause a cooling, and order-of-magnitude estimates suggest that the global impact could be significant⁶². It is difficult to quantify the global significance of this climate forcing, however, because it has only been measured for extended stagnant conditions and global aerosol data are not available.

Anthropogenic aerosol/cloud climate forcing must also occur over continents, even though natural aerosol amounts are larger there. Cases of decreased cloud reflectivity have been observed in polluted industrial regions⁶³, but such cases seem to be the exception, not the rule. Large absorbing aerosols (such as carbonaceous ones) are common close to some sources, but the most extensive long-lived anthropogenic aerosols seem to be highly reflective sulphates formed from SO_2 . Thus, in regions such as the continental United States, which has an extensive network of SO_2 sources, we can speculate about the possibility of significant aerosol or aerosol–cloud cooling. There is evidence for increased clouds over the United States^{64,65}, and the warming of $0.2\text{--}0.3\text{ }^\circ\text{C}$ there in the past century^{66,67} is less than the global warming of $0.5\text{--}0.6\text{ }^\circ\text{C}$ ^{50,51}. This is weak circumstantial evidence, however, because the deviation of the US temperature trend from the global trend is in the range of natural variability for an area covering only 1.5% of the Earth.

We conclude that the lack of global aerosol data makes it impossible at present to determine the net anthropogenic aerosol forcing of the climate system. Observation of similar Northern and Southern Hemisphere warmings during the past century^{50,51}, although most anthropogenic aerosols were added in the Northern Hemisphere, is consistent with the expectation that greenhouse forcing dominates over aerosol forcing. But, as Wigley⁶⁸ has emphasized, there is sufficient observational and statistical uncertainty in the hemispheric temperature trends to mask a substantial aerosol forcing, as much as about half of the greenhouse forcing. Indeed, Durkee⁶⁹ has found evidence in satellite data of interhemispheric differences of aerosol and cloud properties, consistent with larger aerosol and aerosol–cloud effects in the Northern Hemisphere. Despite hemispheric dissimilarities in land cover and natural aerosols, such satellite data should eventually help to quantify the effect of aerosols.

Satisfactory quantitative analysis of the net climate forcing

owing to anthropogenic aerosols will be difficult, because of the inhomogeneous distribution of the aerosols. It will be necessary to monitor global tropospheric aerosol properties, and carry out *in situ* case studies under a broad variety of conditions. Such data could yield the direct aerosol climate forcing, and, in conjunction with global cloud data, it could also allow evaluation of aerosol–cloud interactions.

Implications

Climate forcings. Anthropogenic greenhouse gases have increased steadily during the past century, and now cause a global climate forcing of $2\text{--}2.5\text{ W m}^{-2}$. More than half of this forcing has been added in the past three decades. If emissions of these gases continue at the present or increased rates, greenhouse forcing will reach a level of at least $4\text{--}5\text{ W m}^{-2}$ by the middle of next century.

Solar variability may also be an important climate forcing mechanism, but solar irradiance would need to decline by 2% to counteract greenhouse forcing of $4\text{--}5\text{ W m}^{-2}$. There is no evidence indicating such a decrease; in fact, ground-based observations of solar irradiance and of the solar diameter suggest an upper limit of several tenths of one per cent for solar changes in the past 275 years. There remains the possibility that small solar changes trigger larger climate forcings; speculations have included the influence of changes in the flux of ultraviolet light on ozone or the effect of energetic solar particles on atmospheric ionization and thus cloud nucleation. But no evidence has been found for a significant impact of these mechanisms on global surface temperature. Given available scientific evidence, it would be foolish to base greenhouse policy on the hope that solar variability will somehow counteract greenhouse warming.

Aerosols are an important climate forcing. A clustering of large volcanic eruptions could counter greenhouse forcing for years or even decades. But at present we cannot forecast overall volcanic activity and have no basis for anticipating a systematic long-term impact of volcanoes on net climate forcing. On the other hand, anthropogenic tropospheric aerosols constitute a significant climate forcing, due to both their direct radiative effect and their influence on cloud optical properties. Clearly, the tropospheric aerosol forcing is one of cooling, but its magnitude is very uncertain. In the absence of appropriate measurements, it is difficult to quantify the extent to which it counteracts greenhouse forcing.

Are there other important global climate forcings? By considering the Earth as a planet, specifically those factors that influence absorption of solar energy and emission of thermal

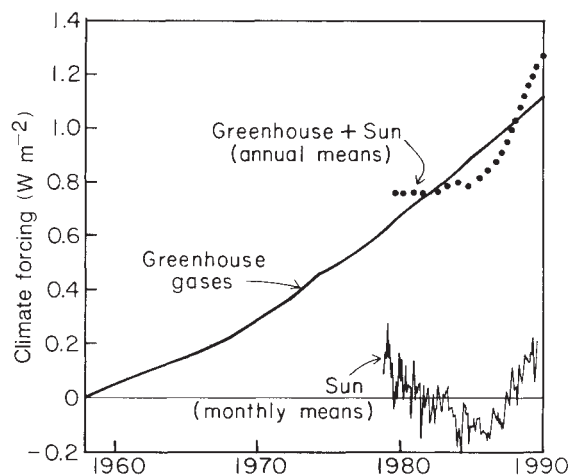
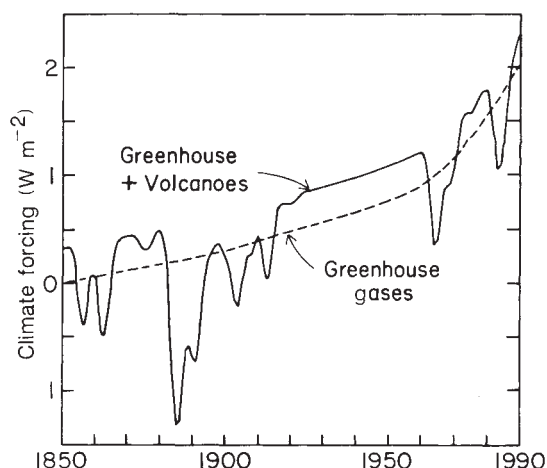


FIG. 3 Climate forcings in past three decades owing to measured changes of greenhouse gases and solar irradiance. Solar irradiance, illustrated for Nimbus 7 data²⁰, has been accurately measured only since 1978. Zero point of solar forcing is the 1978–1989 mean.

FIG. 4 Climate forcings in the past century owing to changes of greenhouse gases and stratospheric aerosols. Aerosol forcing after 1883 is based on atmospheric transmission measurements; for 1850–1883 the three largest volcanoes identified by Lamb⁴⁴ are included, with optical depth scaled relative to Agung, in proportion to the volume of ejecta. Aerosol optical depth for 1883–1960⁴¹ is based on atmospheric transmission measurements at astronomical observatories; subsequently it is based on the transmission of sunlight through the stratosphere as recorded by approximately annual lunar eclipses^{42,43}. Zero point of aerosol forcing is the 1850–1989 mean.



energy, it becomes apparent that one remaining possibility is change of reflectivity of the Earth's surface. Both deforestation and desertification tend to increase planetary reflectivity, but quantitative studies suggest that the climate impacts are confined mainly to the regions of surface change⁷⁰.

Climate response. We have so far restricted our comparisons to climate forcings thus avoiding the uncertainty about global climate sensitivity and the effect of response time. But substantial climate change must be anticipated for the forcings identified. Climate models^{4–6} suggest an equilibrium global warming of 1.5–5.5 °C for doubled CO₂, which is a forcing of 4–4.5 W m⁻². The lower end of this range represents little net feedback, that is, it is approximately the blackbody temperature increase required to yield a flux of ~4 W m⁻². The upper end represents a net positive feedback of about a factor of four. This large range arises because, although feedbacks such as water vapour, sea ice and snow cover are qualitatively understood, even the direction of others such as cloud⁷¹ and aerosol–cloud feedbacks⁷², are uncertain. Recent climate model studies with alternative cloud formulations⁷³ reemphasize the importance of cloud feedbacks, but do not yield a climate sensitivity outside the above range. A warming of even 1.5 °C would make the Earth hotter than it has been in hundreds of thousands of years.

Lindzen^{74,75} has argued that climate sensitivity may be much less than 1.5 °C for doubled CO₂. His argument, buttressed by a metaphysical presumption of a need of the planet for stability, is based primarily on the hypothesis that the net impact of increased moist convection (expected to accompany greenhouse heating) is a drying of the atmosphere, resulting in a negative water–vapour feedback. His descriptive model of moist convection includes the effect of subsidence of surrounding air, but excludes the effects of moisture detrainment at the cloud tops, cirrus anvils, large-scale dynamics and other processes. Lindzen's conclusion can be disproved on several grounds. For example, his proposed moist-convection mechanism can be tested by comparing the real atmosphere in winter and summer: at all latitudes the large-scale impact of increased summer convection is to moisten, not dry, the upper troposphere⁷⁶. Also, satellite measurements show that even the geographical variation of the greenhouse effect⁷⁷ is in close agreement with water–vapour variations and corresponding radiative calculations, and that the maximum greenhouse effect is in regions of intense moist convection. Thus Lindzen's hypothesis of a negative water–vapour feedback cannot be reconciled with real world data, a conclusion that does not depend on uncertainties in global climate model simulations.

The best empirical measure of global climate sensitivity is provided by comparing the Earth's radiation balance in glacial and interglacial conditions. Although glacial–interglacial climate swings are associated with small changes in the Earth's

orbit⁷⁸, these orbital fluctuations have little direct effect on the planetary radiation balance. Instead, the glacial–interglacial temperature differences are maintained by large changes in factors such as planetary surface reflectivity and atmospheric composition. These factors may be climate feedbacks on timescales of millennia, but on timescales of a few years or decades they are climate forcings or boundary conditions. Comparison of the glacial–interglacial global temperature change with the forcing owing to the changed boundary conditions thus provides a measure of climate sensitivity, including all fast feedback processes⁸. The last ice age, which peaked ~20,000 years ago, was 5 ± 1 °C colder than the present interglacial period, and the global climate forcing owing to changed boundary conditions (mainly changes of ice-sheet area, vegetation cover, atmospheric CO₂, CH₄ and aerosols) is estimated^{8,79} to have been 7 ± 2 W m⁻². The inferred climate sensitivity is thus 3 °C per 4 W m⁻² forcing. But the uncertainties in temperature and climate forcing allow a range 2–5 °C per 4 W m⁻² forcing, a result remarkably similar to the range 1.5–5.5 °C estimated on strictly theoretical grounds. This empirical result includes not only water vapour, cloud, snow and sea-ice feedbacks, but any other feedbacks that exist in the real world, such as the aerosol–cloud feedback suggested by Shaw⁸⁰ and Charlson *et al.*⁷².

One consequence of the uncertainty in climate sensitivity is an even larger uncertainty in the response time of the climate system. If climate sensitivity is 1.5–2 °C for a forcing of 4 W m⁻², most of the warming is realized within one to two decades after the forcing is applied; but if the sensitivity is 4–5 °C, the response time may be more than a century^{8,22,23}. The climate response time increases strongly with increasing climate sensitivity, more rapidly than linearly²², because the positive climate feedbacks associated with a high sensitivity only come into play in response to the warming. Because most of the greenhouse forcing has been added very recently (Figs 2, 4), it is clear that a large part of the eventual greenhouse warming owing to anthropogenic gases already in the atmosphere has not yet occurred. This unrealized warming calls into question a 'wait and see' policy towards the greenhouse issue, because the magnitude of this climate time-bomb will grow if greenhouse-gas emissions continue to increase.

Measurements. The interest in global climate change raises the possibility of support for the measurements of climate forcings, feedbacks and diagnostic parameters, which will be needed to sort out cause and effect of climate change. The above discussion emphasizes the need for data on all significant competing factors, including solar irradiance and atmospheric aerosols.

The most precise measurements of solar irradiance recently ended as the Solar Maximum Mission was brought down by atmospheric drag; plans for the Space Shuttle to boost the Solar Maximum Mission to higher orbit were cancelled in the wake

of the Challenger disaster and the resulting reduction in mission frequency. Hopefully, although the measurements are less precise, solar instruments on Nimbus 7 and Earth Radiation Budget satellites will continue to function until the planned 1992 launch of the Upper Atmospheric Research Satellite, with its active-cavity radiometer.

The difficulty of achieving adequate precision in solar-irradiance measurements is highlighted by the changes in the Nimbus 7 data that occur when corrections are made for small errors in the assumed instrument pointing¹¹. The precision needed to reveal decadal trends in the solar irradiance can be provided by instruments whose sensors are internally monitored for degradation of sensitivity, as was the case for the instrument of the Solar Maximum Mission, provided there is temporal overlap of successive instruments. Because the time of failure of an instrument or satellite is not generally predictable, and because the instruments are not extremely expensive, the approach should be to have two solar instruments simultaneously in orbit, thus providing a valuable cross-check as well as continuity of calibration.

Measurements of changes of the solar diameter with the accuracy achievable from space are also important. In conjunction with the irradiance data, these will serve to clarify the relation of diameter and luminosity, and perhaps aid eventual predictability of long-term solar changes.

Aerosols are the source of our greatest uncertainty about climate forcing. Tropospheric aerosols are difficult to monitor because of their spatial inhomogeneity, but they are a crucial variable because of the strong anthropogenic influence on their amount. Not only will it be necessary to monitor the aerosols, but also to have continued global cloud observations, because of possible interactions between aerosols and clouds.

The concept of a 'mission to planet Earth', involving satellites from the United States, Europe and Japan, is presently undergoing intense study and early development. Although intended to cover many aspects of global change, this project has potential for important contributions to the understanding of climate change, provided care is taken that it does not displace resources from other high-priority climate research and monitoring. In planning these activities it is important to recognize that a crucial requirement for climate data sets is long-term continuity. Related to this is the fact that many data sets are best obtained from a variety of conventional sources: operational weather satellites, ships, the world weather network, ground-based and aircraft special studies, and low-cost small-satellite missions. A new initiative could provide a great service by providing resources for analysis of presently available data, and by identifying and filling key gaps in the data.

Policy. In view of the enormous economic and social implications of climate change, researchers must communicate with policy makers, and, indeed, some useful advice can now be given. Based on the complexity of the scientific issues about climate forcings (such as tropospheric aerosols), climate feedbacks (such as clouds) and climate response time (such as ocean mixing and heat storage) it is clear that, contrary to recent advice delivered to the US administration³², the scientific issues will not be settled in 3–5 years. Also, bigger computers, by themselves, will contribute little to our understanding of these problems. What is needed most is support for research using existing data, continuation and improvement of observations, and, perhaps most of all, training of people to define and analyse future data.

Nor can researchers presently provide a prescription for how the world could achieve climate stabilization, should policy-makers decide that such a goal is desirable. This is illustrated by a *gedanken* (thought) experiment concerning the rate of the use of fossil fuel. The burning of fossil fuel releases CO₂, causing about half of the anthropogenic greenhouse effect, but it also releases SO₂, forming atmospheric aerosols which partially counter greenhouse warming by reflecting sunlight and increas-

ing cloud cover. The lifetime of added CO₂ is of the order of 10² years (although it cycles through the biosphere more rapidly) whereas the lifetime of sulphate aerosols is only a few days. This difference in lifetimes has important implications for the effectiveness of changes in fossil-fuel use.

First, as an extreme labelled Case I, we assume that tropospheric aerosols generated by fossil fuels have cancelled a large fraction of the anthropogenic greenhouse effect, say about half, which is approximately the amount of the greenhouse effect owing to CO₂. Case I is conceivable, because observed warming could have been caused by the other greenhouse gases. In Case I, if fossil-fuel use were stabilized (or reduced) warming would accelerate, because the short-lived anthropogenic aerosols would stabilize (or decrease) but CO₂ would continue to increase. CO₂-induced warming is eliminated in Case I only by continued exponential increase of fossil-fuel use. But that would be a Faustian bargain, because fossil fuels would run out, whereupon a huge CO₂-induced warming would begin.

As Case II, we assume that aerosol cooling is negligible compared to CO₂-induced warming. In that case, any reduction in the use of fossil fuel reduces the growth of CO₂ warming in a straightforward way. The real world almost certainly lies somewhere in the continuum between Cases I and II, but we do not know where. Thus, until the research on aerosols is carried out (a difficult task), we do not even know the direction of the change of climate forcing on decadal timescales which would be caused by a modification of fossil-fuel use.

In the presence of such uncertainties, can any useful advice be given to policy makers? We believe so. It is clearly desirable to reduce the ultimate magnitude of the 'experiment' that man is carrying out on the Earth, and there are many actions that could accomplish that and which make good sense on other grounds. These include phasing out CFCs, improving energy efficiency, increasing recycling, reducing deforestation and planting trees in appropriate places. All of these are needed for other reasons, and, in the long-term, pay for themselves. It will, however, require unprecedented international cooperation to achieve a turnaround of global greenhouse-gas emissions. The Montreal Protocol, motivated by a common desire to protect the stratospheric ozone layer, promises significant progress on CFC reductions. The first steps of an analogous process for climate protection are being taken by the Intergovernmental Panel on Climate Change, but it is difficult to predict its prospects for success.

Finally, we would be remiss if, in discussing the policy implications of anthropogenic climate change, we did not raise the issues of alternative energy sources and global population. It seems imperative that governments give much higher priority to research and development on energy sources that produce little or no greenhouse gases. Otherwise we risk the danger of soon finding ourselves, in the mid-American vernacular, up the proverbial creek without a paddle. Also, it is obvious that even sizable reduction of *per capita* greenhouse-gas emissions will be negated if global population continues to increase at the present rate. Thus governments must foster conditions leading to population stabilization, if we are to preserve the global climate and environment. □

The authors are at the NASA Goddard Institute for Space Studies, 2880 Broadway, New York, New York 10025, USA.

1. Keeling, C. D., Bacastow, R. & Whorf, T. *Carbon Dioxide Review* (ed. Clark, W. C.) 375–385 (Clarendon, New York, 1983).
2. Barnola, J. M. et al. *Nature* **329**, 408–414 (1987).
3. Weiss, R. F. *J. geophys. Res.* **86**, 7185–7195 (1981).
4. Charney, J. *Carbon Dioxide and Climate* (Natl. Acad. Sci., Washington, DC, 1979).
5. Hansen, J. et al. *Science* **213**, 957–966 (1981).
6. Dickinson, R. E. & Cicerone, R. J. *Nature* **319**, 109–115 (1986).
7. Manabe, S. & Wetherald, R. T. *J. Atmos. Sci.* **32**, 3–15 (1975).
8. Hansen, J. et al. *Geophys. Monogr. Ser.* **29**, 130–163 (1984).
9. Hansen, J. et al. *J. geophys. Res.* **93**, 9341–9364 (1988).

10. Wigley, T. M. L. & Raper, S. C. B. *Nature* **344**, 324–327 (1990).
11. Hansen, J., Lacis, J. & Ruedy, R. in *Climate Impact of Solar Variability*, Conf. Publ. 3086, 149–159 (NASA, Greenbelt, 1990).
12. Rasmusson, E. M. *Am. Sci.* **73**, 168–178 (1985).
13. Dickinson, R. E. *The Greenhouse Effect Climatic Change and Ecosystems*, 206–270 (Wiley, New York, 1986).
14. Mitchell, J. F. B. *et al. Scientific Assessment of Climate Change* Section 5 (World Meteorological Organisation/United Nations Environment Programme, Geneva, in the press).
15. Hansen, J., Lacis, J. & Prather, M. *J. geophys. Res.* **94**, 16417–16421 (1989).
16. Rowland, F. S. & Isaksen, I. S. A. *The Changing Atmosphere* (Wiley, New York, 1988).
17. Lacis, A. A., Wuebbles, D. J. & Logan, J. A. *J. geophys. Res.* **95**, 9971–9981 (1990).
18. Shine, K. *et al. Scientific Assessment of Climate Change* Section 2 (World Meteorological Organisation/United Nations Environment Programme, Geneva, in the press).
19. Luther, F. M. & Fouquart, Y. *Intercomparison of Radiation Codes in Climate Models*, *World Climate Prog. Rep.* 93 (World Meteorological Organisation, Geneva, 1984).
20. Hickey, J. R., Alton, B. M., Kyle, H. L. & Hoyt, D. *Space Sci. Rev.* **48**, 321–342 (1988).
21. Willson, R. C. & Hudson, H. S. *Nature* **332**, 810–812 (1988).
22. Hansen, J. *et al. Science* **229**, 857–859 (1985).
23. Wigley, T. M. L. & Schlesinger, M. E. *Nature* **315**, 649–652 (1985).
24. Harvey, L. D. D. & Schneider, S. H. *J. geophys. Res.* **90**, 2192–2205 (1985).
25. Lean, J. *Science* **244**, 197–200 (1989).
26. Wetherald, R. T. & Manabe, S. *J. Atmos. Sci.* **32**, 2044–2059 (1975).
27. Sofia, S., Demarque, P. & Endal, A. *Am. Sci.* **73**, 326–333 (1985).
28. Foukal, P. & Lean, J. *Science* **247**, 556–558 (1990).
29. Gough, D. *The Solar Output and its Variations* 451–474 (Colorado Univ. Press, Boulder, 1977).
30. Eddy, J. A. *Science* **192**, 1189–1202 (1976).
31. Wigley, T. M. L. & Kelly, P. M. *Phil. Trans. R. Soc. A* **330**, 547–560 (1990).
32. Sert, F., Jastrow, R. & Nierenberg, W. A. *Scientific Perspectives on the Greenhouse Problem* (Marshall Institute, Washington, DC, 1989).
33. Radick, R. R., Lockwood, G. W. & Ballunas, S. L. *Science* **247**, 39–44 (1990).
34. Hoyt, D. V. *Rev. Geophys. Space Phys.* **17**, 429–458 (1979).
35. Gilliland, R. L. *Astrophys. J.* **248**, 1144–1155 (1981).
36. Gilliland, R. L. *Astrophys. J.* **253**, 399–405 (1982).
37. Endal, A. S., Sofia, S. & Twigg, L. W. *Astrophys. J.* **290**, 748–757 (1985).
38. Damon, P. E. & Sonett, C. P. in *The Sun in Time* (Univ. Arizona Press, Tucson, in the press).
39. Stuiver, M. & Braziunas, T. F. *Nature* **338**, 405–408 (1989).
40. Sparks, J. *Life of Benjamin Franklin* Vol. 6, 455–457 (Tappan & Dennett, Boston, 1844).
41. Pollack, J. B. *et al. Nature* **263**, 551–555 (1976).
42. Keen, R. A. *Science* **222**, 1011–1013 (1983).
43. Hansen, J. E. & Matsushima, S. *J. geophys. Res.* **71**, 1073–1081 (1966).
44. Lamb, H. H. *Phil. Trans. R. Soc. A* **255**, 425–533 (1970).
45. Rasool, S. I. & Schneider, S. H. *Science* **173**, 138–141 (1971).
46. Bryson, R. A. *Science* **184**, 753–760 (1974).
47. Shaw, G. E. & Khalil, M. A. K. *Handbook of Environmental Chemistry* (ed. Hutzinger, O.) (Springer, Berlin, 1989).
48. MacCracken, M. C., Cess, R. D. & Potter, G. L. *J. geophys. Res.* **90**, 14445–14450 (1985).
49. Hansen, J. E., Lacis, A. A., Lee, P. & Wang, W. C. *Ann. N.Y. Acad. Sci.* **338**, 575–587 (1980).
50. Jones, P. O., Wigley, T. M. L. & Wright, P. B. *Nature* **322**, 430–434 (1986).
51. Hansen, J. & Lebedeff, S. *J. geophys. Res.* **92**, 13345–13372 (1987).
52. Toon, O. B. & Pollack, J. B. *J. appl. Met.* **15**, 225–246 (1976).
53. Coakley, J. A. & Cess, R. D. *J. Atmos. Sci.* **42**, 1677–1692 (1985).
54. Schwartz, S. E. *Nature* **336**, 441–445 (1988); **340**, 515–516 (1989).
55. Twomey, S. *J. Phys. Chem.* **84**, 1459–1463 (1980).
56. Twomey, S. & Wojciechowski, T. A. *J. Atmos. Sci.* **26**, 684–688 (1969).
57. Conover, J. H. *J. Atmos. Sci.* **23**, 778–785 (1966).
58. Hobbs, P. V., Radke, L. F. & Shumway, S. E. *J. Atmos. Sci.* **27**, 81–89 (1970).
59. Coakley, J. A., Bernstein, R. L. & Durkee, P. A. *Science* **237**, 1020–1022 (1987).
60. Radke, L. F., Coakley, J. A. & King, M. D. *Science* **246**, 1146–1149 (1989).
61. Albrecht, B. A. *Science* **245**, 1227–1230 (1989).
62. Twomey, S. A., Peigross, M. & Wolfe, T. L. *Tellus* **B36**, 356–366 (1984).
63. Kondratyev, K., Binenov, V. & Petrenchuk, O. *Atmos. oceanic Phys.* **17**, 122–127 (1981).
64. Angell, J. K., Korshover, J. & Cotton, G. F. *J. Clim. appl. Met.* **23**, 752–761 (1984).
65. Plantico, M. S., Karl, T. R., Kukla, G. & Gavin, J. *J. geophys. Res.* (in the press).
66. Karl, T. R., Baldwin, R. G. & Burgin, M. G. *NOAA Historical Climatology Ser. 4–5* (National Climate Data Center, Asheville, 1988).
67. Hansen, J. *et al. Coping with Climate Change* 68–81 (Climate Institute, Washington, DC, 1989).
68. Wigley, T. M. L. *Nature* **339**, 365–367 (1989).
69. Durkee, P. A. *Atmos. Envir.* (in the press).
70. Henderson-Sellers, A. & Gornitz, V. *Clim. Change* **6**, 231–257 (1984).
71. Cess, R. D. *et al. Science* **245**, 513–516 (1989).
72. Charlson, R. J., Lovelock, J. E., Andreae, M. O. & Warren, S. G. *Nature* **326**, 655–661 (1987).
73. Mitchell, J. F. B., Senior, C. A. & Ingram, W. J. *Nature* **341**, 132–134 (1989).
74. Kerr, R. A. *Science* **246**, 1118–1119 (1989).
75. Lindzen, R. *Bull. Am. met. Soc.* **71**, 288–299 (1990).
76. Oort, A. H. & Rasmusson, E. M. *Prof. Pap. NOAA* **14** (1971).
77. Raval, A. & Ramanathan, V. *Nature* **342**, 758–761 (1989).
78. Hays, J. D., Imbrie, J. & Shackleton, N. J. *Science* **194**, 1121–1132 (1976).
79. Harvey, L. D. *Nature* **334**, 333–335 (1988).
80. Shaw, G. E. *Clim. Change* **5**, 297–303 (1983).

ACKNOWLEDGEMENTS. We thank M. MacCracken and T. Wigley for suggestions, and H. L. Kyle and R. A. Keen for data. This work was supported by NASA and the EPA.

Stimulation of p21^{ras} upon T-cell activation

Julian Downward*, Jonathan D. Graves†, Patricia H. Warne*, Sydonia Rayter* & Doreen A. Cantrell†

*Signal Transduction Laboratory and †Lymphocyte Activation Laboratory, Imperial Cancer Research Fund, Lincoln's Inn Fields, London WC2A 3PX, UK

External signals that control the activity of proteins encoded by the *ras* proto-oncogenes have not previously been characterized. It is now shown that stimulation of the antigen receptor of T lymphocytes causes a rapid activation of p21^{ras}. The mechanism seems to involve a decrease in the activity of GAP, the GTPase-activating protein, on stimulation of protein kinase C. In lymphocytes, p21^{ras} may therefore be an important mediator of the action of protein kinase C.

THE activated *ras* oncogenes (Ha-, Ki- and N-*ras*) can transform mammalian cells in culture and have been implicated in the formation of a high proportion of human tumours¹. The *ras* oncogenes encode closely related proteins of relative molecular mass 21,000 (p21^{ras}) that bind GTP and catalyse its hydrolysis to GDP. Ras proteins can stimulate cellular growth and other events when they are bound to GTP ('active') but not when they are bound to GDP ('inactive')^{2–4}. Mutations that allow p21^{ras} to transform cells all cause accumulation of the GTP-bound form of the protein; many inhibit the intrinsic GTPase activity of p21^{ras} (ref. 5) and others increase the rate of exchange of bound nucleotide with cytoplasmic pools, which contain much more GTP than GDP⁶.

No cellular stimulus has previously been identified that controls the activation state of mammalian p21^{ras}, as measured by the amount of GTP bound to it relative to GDP. But a likely component of the *ras* pathway has been identified in GAP, a protein that stimulates the hydrolysis of GTP on p21^{ras}, thereby causing it to enter the inactive GDP-bound state⁷. Ways in which the activity of GAP might be regulated remain the subject of speculation.

Here we report that p21^{ras} is indeed part of a signal transduction pathway and that in T lymphocytes its activation state can be rapidly regulated by an extracellular stimulus operating on a cell-surface receptor.

Stimulation of p21^{ras}

The activation state of p21^{ras} was measured in intact T cells by metabolic labelling with [³²P]orthophosphate and then lysing them and immunoprecipitating p21^{ras}. Figure 1a is an autoradiogram showing the nucleotides bound to the endogenous p21^{ras} in human peripheral blood lymphoblasts (PBLs)⁸ and in the human T-cell line Jurkat⁹. Although p21^{ras} from untreated cells is almost entirely GDP-bound, on treatment with the T-cell-activating lectin phytohaemagglutinin (PHA)¹⁰ or CD3-specific monoclonal antibody UCHT-1 (ref. 11), a large amount of GTP accumulates on p21^{ras}. In the cell line HPB-ALL, where the antigen receptor is uncoupled from phosphatidylinositol (PtdIns) turnover¹², reagents specific for the antigen receptor fail to influence the nucleotide bound to p21^{ras} (Fig.

FIG. 1 Effect of T-cell activation on the nucleotide bound to p21^{ras}. **a**, Thin layer chromatogram of the nucleotides eluted from immunoprecipitates of p21^{ras} from human peripheral blood lymphoblasts labelled with ³²P-orthophosphate (PBLs, 1–4), J6 Jurkat cells (5–8) and HPBALL cells (9–12). Immunoprecipitation was with monoclonal antibody Y13-259 in all lanes except 1, 5 and 9, where no specific antibody was added. Cells were unstimulated (2, 6, 10) or stimulated for 30 min with 10 µg ml⁻¹ monoclonal antibody UCHT-1 (3, 7, 11) or PHA (1, 4, 5, 8, 9, 12). The positions of GTP and GDP standards are indicated. **b**, Time course showing the nucleotides bound to p21^{ras} on stimulation of PBLs with PHA (○) or UCHT-1 (□). The p21^{ras} was immunoprecipitated with antibody 259 from cells labelled with ³²P-orthophosphate that had been stimulated for the indicated times. Nucleotides were separated by TLC and quantitated by direct scanning for β radiation.

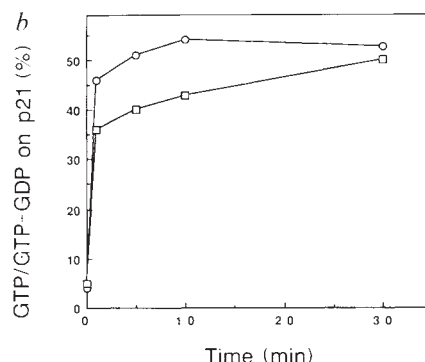
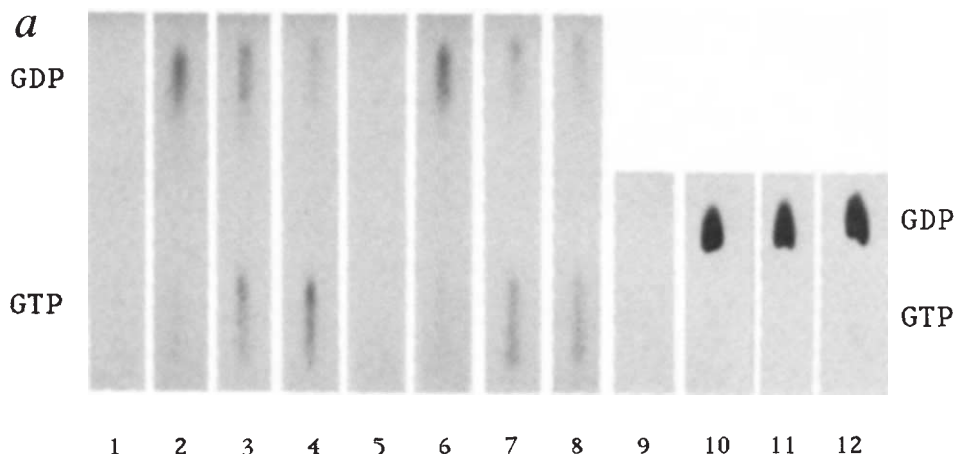
METHODS. Cells were isolated, cultured and labelled as described previously⁸. Lysis was performed in 50 mM HEPES buffer, pH 7.4, 1% Triton X-100, 100 mM NaCl, 5 mM MgCl₂, 1 mg ml⁻¹ BSA, 10 mM benzamidine, 10 µg ml⁻¹ leupeptin, 10 µg ml⁻¹ aprotinin, 10 µg ml⁻¹ soybean trypsin inhibitor. Nuclei were removed by centrifugation at 15,000g for 2 min and 0.5 M NaCl, 0.5% deoxycholate and 0.05% sodium dodecyl sulphate added to the lysate. Immunoprecipitation was for 40 min using antibody 259 precoupled to protein A-agarose by rabbit anti-rat immunoglobulin. All immunoprecipitations were in duplicate with non-specific controls. Immunoprecipitates were washed with 8 × 1 ml of 50 mM HEPES, pH 7.4, 500 mM NaCl, 5 mM MgCl₂, 0.1% Triton X-100, 0.005% SDS and nucleotide eluted with 2 mM EDTA, 2 mM DTT, 0.2% SDS, 0.5 mM GTP, 0.5 mM GDP at 68 °C for 20 min. Separation of eluted nucleotide was on PEI-cellulose plates run in 1.2 M ammonium formate, 0.8 M HCl.

1a). The increases in GTP levels on cell stimulation are rapid: they are clearly visible within one minute and are maximal by 10 minutes (Fig. 1b). The amount of GTP bound to p21^{ras} as a proportion of total nucleotide rises to nearly 50% after treatment with PHA or UCHT-1; in untreated cells the level is ~5%.

Protein kinase C mediates p21^{ras} activation

Stimulation of the antigen receptor in T lymphocytes is known to activate several signalling pathways¹⁰. To determine which of these is involved in the activation of p21^{ras}, downstream components were directly activated by pharmacological agents. Figure 2a shows the effects of the calcium ionophore ionomycin and the protein kinase C activator phorbol dibutyrate (PDBu) on the amount of GTP bound to p21^{ras}. The phorbol ester causes a great stimulation of p21^{ras}, but ionomycin does not affect the nucleotide bound to p21^{ras}. The fact that PDBu causes a greater stimulation of GTP levels on p21^{ras} than do the agents that act on the T-cell antigen receptor indicates that protein kinase C is likely to be a principal intracellular mediator of the activation of p21^{ras} in these cells. This is consistent with the observation that in HPB-ALL cells PDBu stimulates p21^{ras}, whereas PHA and UCHT-1, which in these cells fail to activate PtdIns turnover and hence protein kinase C¹², do not. As short-term treatment with PDBu has no measurable effects on phospholipid metabolism in T cells (J.D.G. and D.A.C., unpublished observations), we believe that the PDBu-dependent effects are the result of activation of protein kinase C and not direct lipid effects on GAP itself.

The action of PDBu is very rapid, with maximal stimulation occurring within one minute, which is faster than with antigen receptor agonists (Fig. 2b). This is again compatible with protein kinase C mediating the effect of antigen receptor on p21^{ras} activation. Stimulatory effects of PDBu on p21^{ras} occur at con-



centrations as low as 5 ng ml⁻¹ (Fig. 2c): this is similar to the level of PDBu needed to activate protein kinase C¹³. Lysates from PDBu-treated or control ³²P-labelled PBLs were mixed with lysates from unlabelled cells (Fig. 2d); treatment of the unlabelled cells has no influence on labelled GTP bound to p21^{ras}, excluding the possibility of a post-lytic artefact.

p21^{ras} activation in permeabilized cells

Two likely mechanisms for the observed activation of p21^{ras} might involve a decrease in GTP hydrolysis on p21^{ras}, possibly due to inactivation of GAP, the GTPase activating protein⁷, or an increase in the rate at which guanine nucleotide exchanged onto and off p21^{ras}. To address these possibilities, we studied the activation state of p21^{ras} in an alternative way using peripheral blood lymphoblasts that had been permeabilized with streptolysin O (ref. 14). When the permeabilized cells are exposed to [³²P]GTP, the labelled nucleotide binds rapidly to p21^{ras}, which can then be immunoprecipitated and analysed. Very little GTP accumulates on p21^{ras} from untreated cells, but significant amounts of GTP are seen on p21^{ras} from cells treated with PHA or PDBu (Fig. 3a). The rate at which total nucleotide binds is rapid and identical whether or not the cells have been treated with PDBu (Fig. 3b, c). Even at 27 °C (where the exchange reaction is slower than at 37 °C) PDBu treatment does not influence the binding rate. It is very striking that the exchange of nucleotide onto p21^{ras} is so rapid in permeabilized lymphoblasts; half-maximal binding is achieved in about one minute at 37 °C, whereas for pure p21^{ras} or p21^{ras} in isolated membrane preparations (either from quiescent or stimulated cells) about 60 minutes would be required to reach the same level of binding in the same buffer and at the same temperature. This indicates that an activity exists in the permeabilized cells that stimulates the exchange of nucleotide on p21^{ras}: the activity is lost or

inactivated in membrane preparations. This exchange factor seems to be constitutively active for both quiescent and stimulated cells. Exchange factors for $p21^{ras}$ have recently been identified¹⁵⁻¹⁷.

As shown in Fig. 3d (37°C) and Fig. 3e (27°C), $p21^{ras}$ is almost totally GDP-bound in the unstimulated cells, but in the cells treated with PDBu significant amounts of GTP accumulate on $p21^{ras}$. As the ratio of total labelled GTP to GDP free in the permeabilized cells is about 1:1 and is not altered by PDBu treatment (data not shown), the difference in the level of GTP bound must be due to a decrease in the hydrolysis of GTP on $p21^{ras}$ in the cells where protein kinase C has been activated.

Protein kinase C inhibits GAP activity

A possible mechanism for this inhibition of the GTPase activity of $p21^{ras}$ could be the inactivation of GAP on stimulation of protein kinase C within the cell. The activity of GAP in a Triton X-100 lysate of PBLs that had been treated for two minutes with PDBu was compared with the GAP activity in lysates from

untreated cells. The ability of a lysate from PDBu-treated cells to induce hydrolysis of GTP on $p21^{ras}$ is six times less than that of a lysate from untreated cells (Fig. 4). Okadaic acid, a potent inhibitor of phosphatases 1 and 2a, was included in the incubation of the cells with or without PDBu and also in the lysis buffer; okadaic acid alone had no effect on GAP activity. As the inhibition of GAP activity by PDBu treatment is seen in whole cell lysates, it seems that the effect is on the activity of GAP but not on its localization in the cell.

Protein kinase C controls N-Ras and Ki-Ras

Protein kinase C phosphorylates $p21^{Ki-ras4B}$, although no effect of this modification on the behaviour of the Ras protein has been reported¹⁸. The stimulation of GTP levels on $p21^{ras}$ as a proportion of total nucleotide is in fact higher when antibody 259, which recognizes all Ras proteins, is used than when antibody 238, which recognizes only $p21^{Ha-ras}$ and $p21^{Ki-ras}$, is used (Fig. 5a). Reimmunoprecipitation of the supernatants remaining after immunoprecipitation with antibody 238 shows that $p21^{N-ras}$

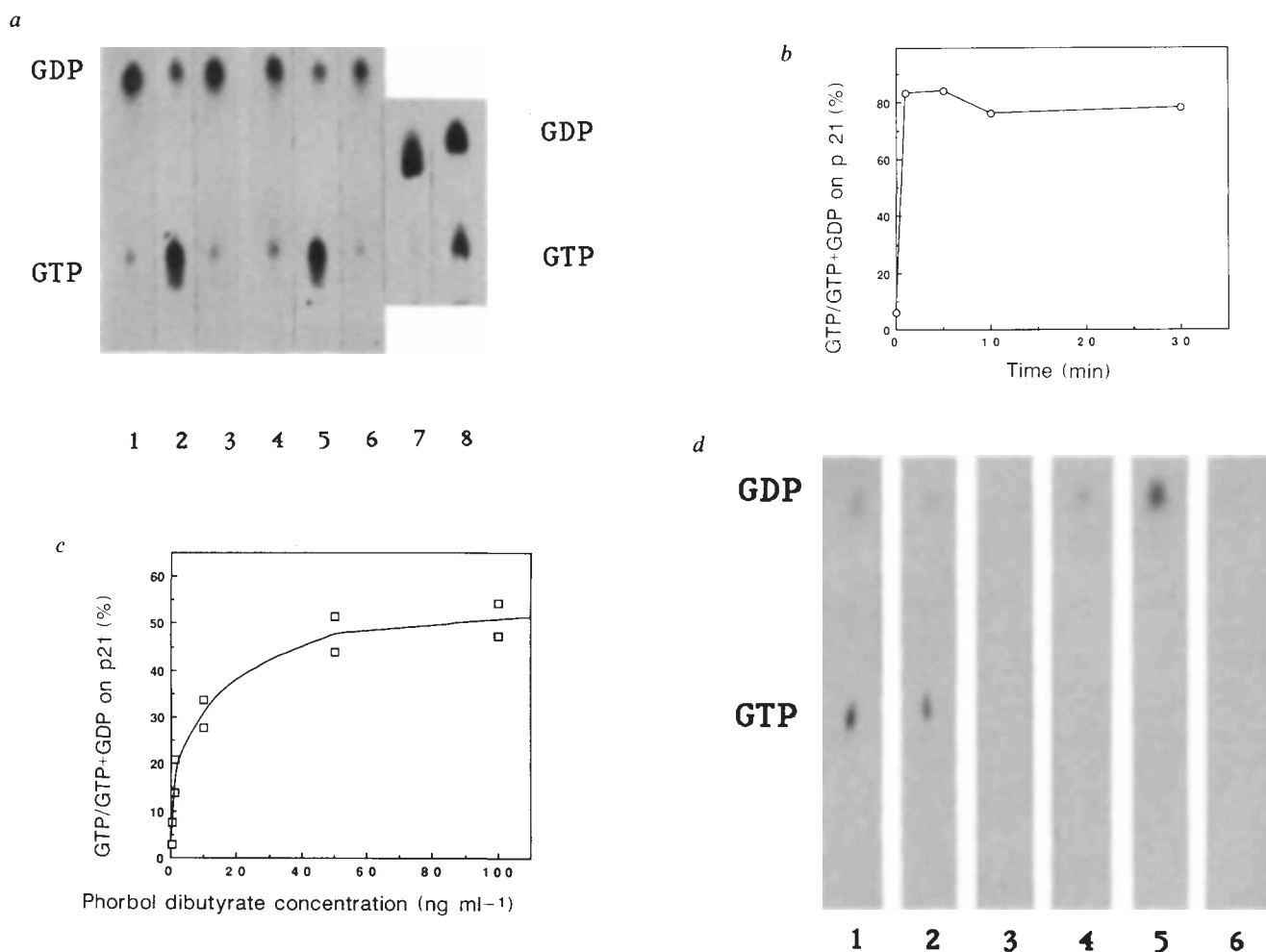


FIG. 2 Effect of phorbol ester treatment on the activation state of $p21^{ras}$. **a**, Thin layer chromatogram of the nucleotides eluted from immunoprecipitates of $p21^{ras}$ from human PBLs labelled with ^{32}P -orthophosphate (1, 2, 3), Jurkat cells (4, 5, 6) and HPB-ALL cells (7, 8). Immunoprecipitation was with monoclonal antibody Y13-259. Cells were unstimulated (1, 4, 7) or stimulated for 30 min with 100 ng ml⁻¹ PDBu (2, 5, 8) or 5 μ g ml⁻¹ ionomycin (3, 6). The position at which GTP and GDP standards ran is indicated. **b**, Time course showing the nucleotides bound to $p21^{ras}$ on stimulation of PBLs with 100 ng ml⁻¹ PDBu. The $p21^{ras}$ was immunoprecipitated with antibody 259 from cells labelled with ^{32}P -orthophosphate that had been stimulated for the indicated times. Nucleotides were separated by TLC and quantitated by direct scanning for β radiation. **c**, Phorbol dibutyrate concentration depen-

dence of the nucleotides bound to $p21^{ras}$ in PBLs. The $p21^{ras}$ was immunoprecipitated with antibody 259 from cells labelled with ^{32}P -orthophosphate that had been stimulated for 30 min with the indicated concentration. Nucleotides were separated by TLC and quantitated by direct scanning for β radiation. **d**, Thin layer chromatogram of the nucleotides eluted from immunoprecipitates of $p21^{ras}$ from human PBLs labelled with ^{32}P -orthophosphate that had been treated with 100 ng ml⁻¹ PDBu for 30 min before lysis (1, 2, 3) or were untreated (4, 5, 6). Immunoprecipitation was with antibody 259 (1, 2, 4, 5) or non-specific antibody (3, 6). Immediately after lysis the cells were mixed with lysates from identically prepared unlabelled cells that had been treated with PDBu (1, 3, 4, 6) or were untreated (2, 5).

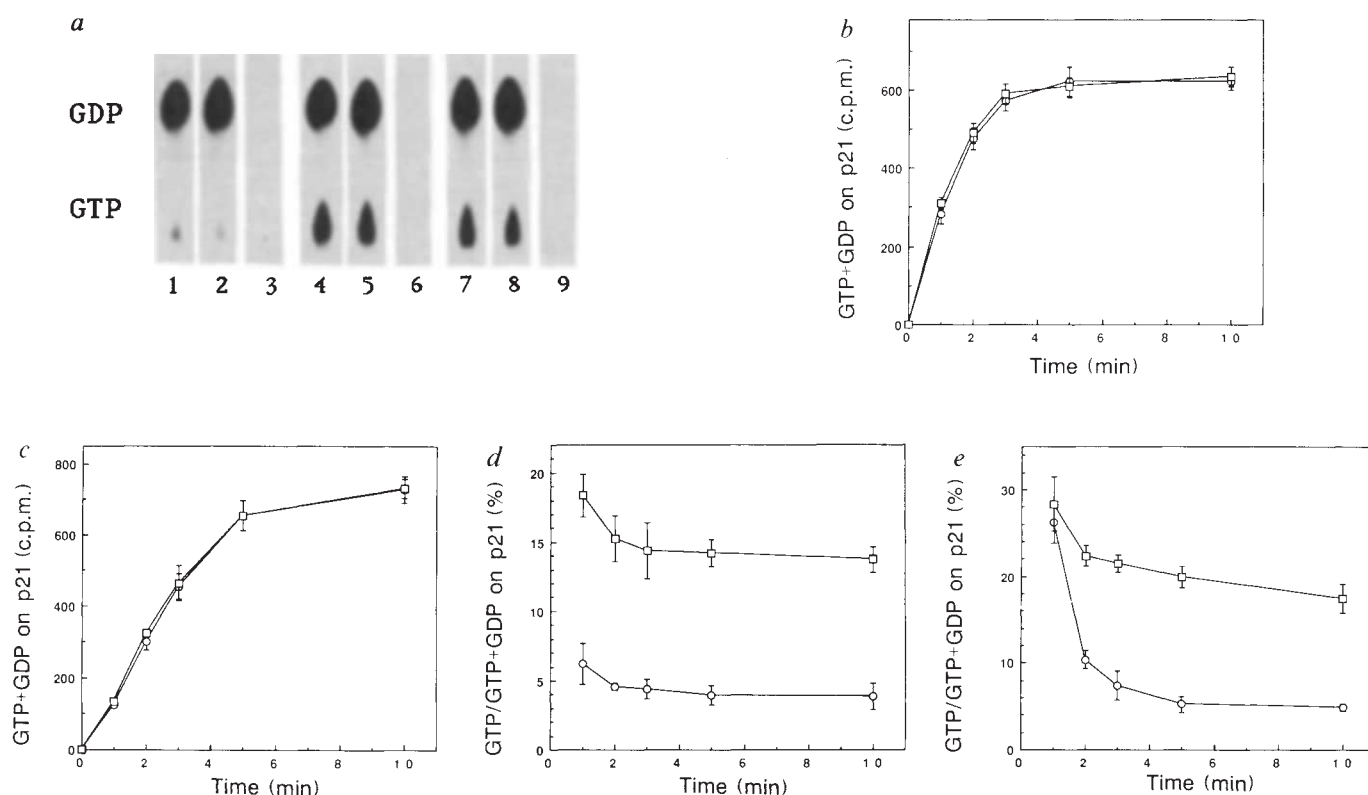
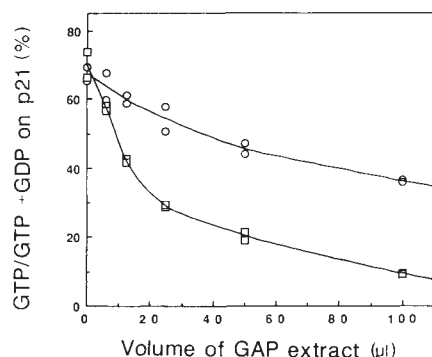


FIG. 3 Effect of antigen receptor or protein kinase C activation on nucleotide binding to p21^{ras} in permeabilized cells. *a*, Thin layer chromatogram of the nucleotides eluted from immunoprecipitates of p21^{ras} from permeabilized human PBLs that had been exposed to [α-³²P]GTP for 5 min. The cells were quiescent (1, 2, 3), treated with 10 μg ml⁻¹ PHA (4, 5, 6) or 100 ng ml⁻¹ PDBu (7, 8, 9). Immunoprecipitation was with monoclonal antibody Y13-259 in all lanes except 3, 6 and 9, where no specific antibody was added. GTP and GDP standards are indicated. *b*, Time course of total α-³²P-labelled nucleotide bound to p21^{ras} in permeabilized PBLs treated with 100 ng ml⁻¹ PDBu (□) or quiescent (○) at 37 °C. *c*, Time course as *b* but at 27 °C. *d*, Time course of the amount of [α-³²P]GTP as a percentage of total α-³²P-

labelled nucleotide bound to p21^{ras} in permeabilized PBLs treated with 100 ng ml⁻¹ PDBu (□) or quiescent (○) at 37 °C. *e*, Time course as *d* but at 27 °C. The level of nucleotide binding seen is consistent with all p21^{ras} in the cells associating with labelled nucleotide.

METHODS. Human PBLs were permeabilized¹⁴ by addition of 0.4 IU ml⁻¹ of streptolysin O in 10 mM PIPES buffer, pH 7.2, containing 120 mM KCl, 30 mM NaCl, 1 mM ATP, 5 mM free Mg²⁺, 100 nM free Ca²⁺. [α-³²P]GTP (5 μCi, 3,000 Ci mmol⁻¹) was used for each assay point of 0.5 ml. Streptolysin O, [α-³²P]GTP and PDBu or lectin were added simultaneously. The p21^{ras} was recovered at the end of the incubation as described in Fig. 1.

alone also has a large amount of GTP bound to it after PDBu stimulation. Immunoprecipitation of p21^{ras} from lysates of [³⁵S]methionine-labelled PBLs (Fig. 5*b*) indicates that these cells contain similar quantities of p21^{N-ras} and p21^{Ki-ras4B}, but little p21^{Ha-ras} or p21^{Ki-ras4A}. Protein kinase C stimulation therefore has at least as great an effect on the activity of the N-ras protein as it does on the Ki-ras protein.



Discussion

The data presented show that in T cells the activation state of p21^{ras} is controlled by the antigen receptor. To our knowledge, this is the first time that a cellular stimulus has been shown to control the activation state of mammalian Ras proteins. The mechanism by which p21^{ras} is stimulated on T-cell activation seems to involve the protein kinase C inhibition of a GTPase

FIG. 4 GAP activity in lysates of peripheral blood lymphoblasts: effect of phorbol ester treatment. Lysates were made from 5×10^8 PBLs that were either quiescent (□) or treated with 100 ng ml⁻¹ PDBu (○) for 2 min and assayed for GAP activity at a range of dilutions.

METHODS. Okadaic acid (100 nM) was included in the 2-min PDBu and control incubations. Lysates were made by disrupting cells in 0.5 ml 10 mM PIPES buffer, pH 7.2, containing 120 mM KCl, 30 mM NaCl, 5 mM free Mg²⁺, 100 nM free Ca²⁺, 1% Triton X-100, 10% glycerol, 100 nM okadaic acid. The lysate was centrifuged at 300,000g for 10 min and the supernatants assayed for GAP activity. Pure bacterially expressed wild-type human p21^{c-Ha-ras} (50 ng) was allowed to bind 25 μCi of [α-³²P]GTP in the presence of 5 mM EDTA at 37 °C for 5 min. then 10 mM MgCl₂, 1 mM GTP and 1 mM GDP were added and 0.75 μCi aliquots added to varying amounts of GAP extracts made up to the same volume with homogenization buffer. The mixtures were incubated for 5 min at 37 °C, the reactions stopped by addition of 10 volumes of cold lysis buffer containing 0.5 M NaCl and p21^{ras} immunoprecipitated using antibody 259. The proportion of [α-³²P]GTP to total labelled nucleotide on p21^{ras} was calculated following separation of GTP from GDP by TLC and quantitation by direct scanning for β-radiation.

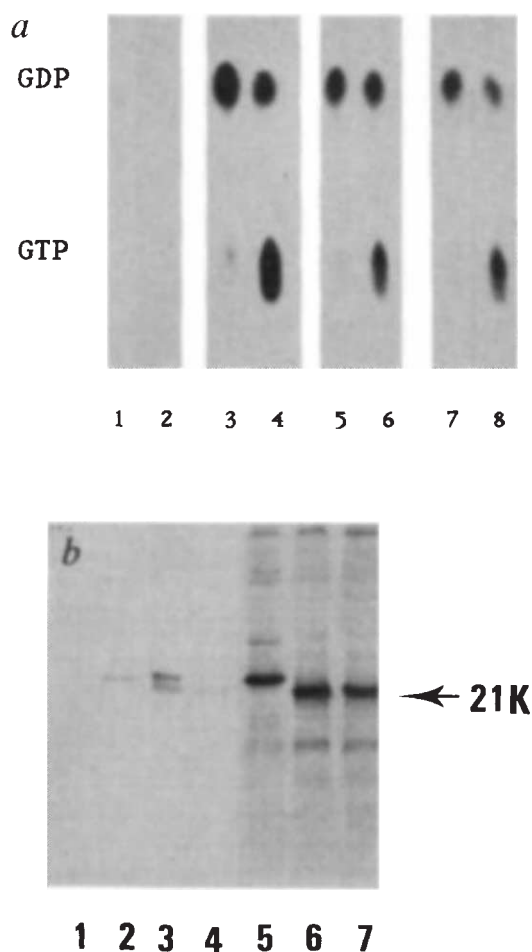


FIG. 5 Effect of protein kinase C activation on the different types of $p21^{ras}$ in peripheral blood lymphoblasts. *a*, Thin layer chromatogram of the nucleotides eluted from immunoprecipitates of $p21^{ras}$ from human PBLs labelled with [^{32}P]orthophosphate that were unstimulated (1, 3, 5, 7) or stimulated for 10 min with 100 ng ml^{-1} PDBu (2, 4, 6, 8). The antibodies used in the immunoprecipitations were nonspecific (1, 2), Y13-259 (3, 4), Y13-238 (5, 6) or Y13-259 on the lysates remaining after immunoprecipitation with Y13-238 (7, 8). *b*, SDS-polyacrylamide gel of immunoprecipitates of $p21^{ras}$ from a lysate of PBLs labelled with [^{35}S]methionine. The antibodies used were non-specific (1), Y13-238 (2), Y13-259 (3), Y13-259 on the lysates remaining after immunoprecipitation with Y13-238 (4). Also shown are Y13-259 immunoprecipitations from the cell line Rat-1.K-ras25 that overexpresses human $p21^{K1-ras4B}$ (5), the cell line Rat-1.H-ras13 that overexpresses human $p21^{Ha-ras}$ (6) and the cell line Rat-1.N-ras16 that overexpresses human $p21^{N-ras}$ (7). Cells were labelled with [^{35}S]methionine as described^{21,8} and lysis and immunoprecipitation carried out as for Fig. 1. For the lymphoblasts 2×10^7 c.p.m. were used per immunoprecipitation and 5×10^5 c.p.m. for lane 6 and 2×10^6 c.p.m. for lanes 5 and 7.

activating protein, presumably GAP. The way in which activated protein kinase C brings about inactivation of GAP remains to be characterized. The simplest mechanism would be a direct phosphorylation of GAP by protein kinase C; however, we have failed to find any evidence of phosphorylation of GAP stimulated by phorbol esters in PBLs or Jurkat cells (data not shown). It is perhaps more likely that protein kinase C phosphorylates a different protein that in turn acts as an inhibitor of GAP. The mechanisms described here seem to be quite different from the previously reported tyrosine kinase phosphorylation of GAP^{19,20} whose functional significance remains unclear. Alterations in the levels of tyrosine phosphorylation in response to phorbol ester treatment have not been demonstrated in lymphocytes.

The very rapid nature of the conversion of $p21.GDP$ to $p21.GTP$ in whole cells on activation of protein kinase C indicates that the rate of exchange of nucleotide on $p21^{ras}$ in the

cell must be very fast compared with that on isolated $p21^{ras}$. The data from the permeabilized cell experiments confirm this and further show that the exchange rate is constitutively elevated whatever the growth state of the cells. It therefore seems that in the cell, fresh guanine nucleotide (primarily GTP, which predominates over GDP in the cytoplasm) is rapidly exchanging onto $p21^{ras}$ because of the action of an exchange protein. In the quiescent lymphoblast the high activity of GAP ensures that bound GTP is rapidly hydrolysed and the balance of nucleotide on $p21^{ras}$ remains in favour of GDP. On activation of protein kinase C the activity of GAP is reduced several fold, allowing the exchange protein to predominate and the nucleotide balance to swing rapidly over to GTP. Thus the activation state of $p21^{ras}$ is the result of a dynamic balance between influences of GAP and of the exchange protein: in this system it seems that GAP is the main target for regulation, although it is clearly possible that the exchange protein could also be regulated. □

Received 30 March; accepted 26 July 1990.

- Barbacid, M. A. *Rev. Biochem.* **56**, 779–827 (1987).
- Satoh, T., Endo, M., Nakamura, S. & Kaziro, Y. *FEBS Lett.* **236**, 185–189 (1988).
- Satoh, T., Nakamura, S. & Kaziro, Y. *Molec. cell. Biol.* **7**, 4553–4556 (1987).
- Field, J., Broek, D., Kataoka, T. & Wigler, M. *Molec. cell. Biol.* **7**, 2128–2133 (1987).
- Sweet, R. W. *et al. Nature* **311**, 273–275 (1984).
- Der, C. J., Pan, B. T. & Cooper, G. M. *Molec. cell. Biol.* **6**, 3291–3294 (1986).
- Trahey, M. & McCormick, F. *Science* **238**, 542–545 (1987).
- Cantrell, D. A., Davies, A. A. & Crumpton, M. J. *Proc. natn. Acad. Sci. U.S.A.* **82**, 8158–8162 (1985).
- Gillis, S. & Watson, J. *Immunol. Rev.* **54**, 81–109 (1981).
- Weiss, A. & Imboden, J. B. *Adv. Immun.* **41**, 1–38 (1987).
- Beverly, P. C. L. & Callard, R. E. *Eur. J. Immun.* **11**, 329–334 (1981).
- Brattsand, G., Cantrell, D. A., Ward, S., Ivars, F. & Gulberg, M. *J. Immun.* **144**, 3651–3658 (1990).

- Rodriguez-Pena, A. & Rozengurt, E. *EMBO J.* **5**, 77–83 (1986).
- Graves, J. D., Lucas, S. C., Alexander, D. R. & Cantrell, D. A. *Biochem. J.* **265**, 407–413 (1990).
- Wolfman, A. & Macara, I. *Science* **248**, 67–69 (1990).
- West, M., Kung, H. & Kamata, T. *FEBS Lett.* **259**, 245–248 (1990).
- Downward, J., Riehl, R., Wu, L. & Weinberg, R. A. *Proc. natn. Acad. Sci. U.S.A.* (in the press).
- Ballester, R., Furth, M. E. & Rosen, O. M. *J. Biol. Chem.* **262**, 2688–2695 (1987).
- Molloy, C. J. *et al. Nature* **342**, 711–714 (1989).
- Ellis, C., Moran, M., McCormick, F. & Pawson, T. *Nature* **343**, 377–381 (1990).
- Downward, J., de Gunzburg, J., Riehl, R. & Weinberg, R. A. *Proc. natn. Acad. Sci. U.S.A.* **85**, 5774–5778 (1988).

ACKNOWLEDGEMENTS. We thank A. Aitken for supplying okadaic acid and P. Parker and E. Rozengurt for constructive criticisms.

Presynaptic mechanism for long-term potentiation in the hippocampus

John M. Bekkers & Charles F. Stevens

The Salk Institute, Howard Hughes Medical Institute, 10010 North Torrey Pines Road, La Jolla, California 92037, USA

Experiments analysing the statistical properties of synaptic transmission, before and after the induction of long-term potentiation (LTP), suggest that expression of LTP largely arises in a presynaptic mechanism—an increased probability of transmitter release.

LONG-TERM potentiation (LTP) is a long-lasting use-dependent increase in the efficacy of excitatory synaptic transmission in the brain that is thought to underlie certain forms of learning and memory^{1,2}. Although the trigger for LTP is generally agreed to occur in the postsynaptic cell^{3–8}, the site at which it is expressed is still disputed. Some authors have suggested that the mechanism of LTP is enhanced neurotransmitter release^{9,10}, others that it is increased postsynaptic sensitivity to transmitter^{11,12}, and still others that both may occur, perhaps in sequence¹³. Here we describe experiments analysing the statistical properties of synaptic transmission, before and after the induction of LTP, in both hippocampal cultures and patch-clamped slices.

Quantal analysis

Since the classic work of Del Castillo and Katz¹⁴, the standard way to determine whether a phenomenon is pre- or postsynaptic has been to carry out a quantal analysis¹⁵. Such an analysis provides estimates of a , the amplitude of response to a single quantum of neurotransmitter, p , the probability of release, and N , the number of release sites. Synaptic transmission becomes more efficacious if a , N or p is increased. In standard treatments, an increase in either N or p indicates a presynaptic mechanism, whereas a postsynaptic mechanism, such as greater receptor sensitivity, would produce an increase in a .

Quantal analysis has, however, proved difficult in the central nervous system for three main reasons. First, only rarely can one find a pair of cells that are monosynaptically connected¹⁶. The formalism for quantal analysis can be extended to include some situations in which more than one input to a neuron is stimulated, but such an extension is especially uncertain for classes of synapses where the Del Castillo and Katz statistical mechanism for release has not been confirmed. The second difficulty is that the characterization of the quantal event, the miniature excitatory postsynaptic current (m.e.p.s.c.), is uncertain both because spontaneous m.e.p.s.cs cannot usually be associated with a specific subset of the many synapses on a typical central neuron, and because cable filtering alters the amplitude and time course of m.e.p.s.cs that originate at a distance from the recording site^{15,17}. Finally, the background noise in intracellular recording is frequently so large that m.e.p.s.cs are obscured^{18,19}.

We have circumvented these difficulties by studying neuronal connections in culture, where synaptically connected pairs of cells can be identified and where m.e.p.s.c. release can be evoked at defined locations so that the inherent properties of m.e.p.s.cs can be separated from cable filtering effects. M.e.p.s.cs at synapses in cultures of hippocampal neurons vary considerably in size, and the analysis of this variability has recently been extended in two ways²⁰. First, m.e.p.s.c. amplitude fluctuations have been demonstrated to arise at single boutons (rather than

reflecting the properties of a population of boutons); second, m.e.p.s.cs in hippocampal slices have been shown to exhibit about the same variability in size previously seen in culture²¹. An accurate quantal analysis of synaptic transmission in the hippocampus must take these properties of m.e.p.s.cs into account (see refs 22, 23).

Analysis in dissociated neurons

In a preliminary series of three experiments in culture, we used the nystatin perforated patch technique²⁴ to observe m.e.p.s.cs in a voltage-clamped neuron before and after briefly perfusing the bath with solution containing no Mg^{2+} . Such perfusion causes cells in the dish to fire spontaneous bursts of action potentials repeatedly and should have potentiated many synapses. Assuming LTP had been induced in a significant number of the neuron's synapses, we should have observed an increase in the mean size of m.e.p.s.cs if expression of LTP is postsynaptic. No detectable change in m.e.p.s.c. size was found up to 96 min after the induction of massive activity in the culture, suggesting that expression of LTP is not postsynaptic. The difficulty with this experiment was, of course, that we could not be sure that the synapses whose m.e.p.s.cs we recorded had undergone LTP. Therefore, we turned to a more conventional protocol.

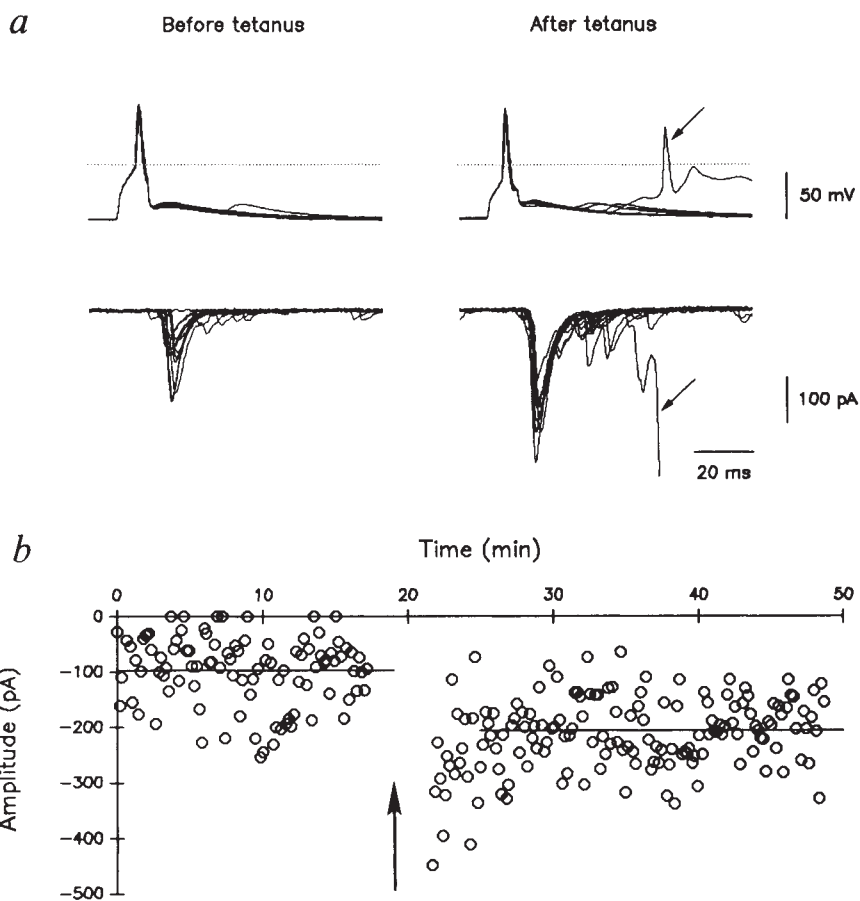
At the outset we failed to observe LTP in 42 neuron pairs when standard whole cell recording methods were used²⁵, and therefore suspected that some critical cytoplasmic factor was being washed out of the cells. Accordingly, all our later experiments used the nystatin perforated patch technique to record from both pre- and postsynaptic neurons.

Figure 1 shows an experiment in which an action potential was repetitively fired (at 0.1 Hz) in one neuron, and the resulting synaptic current was recorded under voltage clamp in a neighbouring neuron (panel *a*). Panel *b* plots the amplitude of the current as a function of time during the experiment. For the first series of stimuli (before the arrow) the mean peak conductance was 1.62 nS, the standard deviation was 1.04 nS and on 8 out of 97 occasions the presynaptic action potential failed to evoke a postsynaptic response. At the arrow the bathing medium containing 10 mM Mg^{2+} (to prevent any possible potentiation during the test stimulation) was exchanged for one with no added Mg^{2+} , the postsynaptic neuron was switched from voltage clamp to current clamp, and the presynaptic neuron was tetanized three times for 2 s at 20 Hz. After this stimulation, the bathing solution with 10 mM Mg^{2+} was reintroduced, postsynaptic voltage clamp was re-established, and the 0.1 Hz stimulation of the presynaptic cell resumed. Following the tetanic stimulation, the mean synaptic current amplitude was 3.43 nS, the standard deviation 0.97 nS, and no failures in transmission occurred in 128 trials. The potentiation factor f in this case was $(3.43/1.62) = 2.1$. The amplitude histograms for pre- and post-tetanus synaptic current are shown in Fig. 2.

The following results argue in favour of the phenomenon in Fig. 1 being LTP. (1) The potentiation was long-lasting. For example, the e.p.s.c. in Fig. 1 was still at the same, elevated level 60 min after the tetanus. In two other cells the roughly threefold potentiation lasted for 80 and 90 min, respectively (at which point the cells were lost). (2) Potentiation could only be induced in a low Mg^{2+} concentration, as expected from the triggering characteristics of LTP^{7,8}. (3) Potentiation was not an

FIG. 1 Long-term potentiation of excitatory synaptic transmission between a pair of rat hippocampal neurons in dissociated cell culture. *a*, Superimposed action potentials recorded in the presynaptic cell (upper traces) and corresponding e.p.s.c.s recorded simultaneously in the postsynaptic cell (lower traces). Left-hand panels were recorded shortly before, right-hand panels about 10 min after, a series of three 2-s tetani at 20 Hz separated by 30 s, performed with the postsynaptic cell in current clamp. All pre- and post-tetanus recordings were made with 10 mM Mg^{2+} solution (see below) in the bath; the tetani were delivered after briefly changing to a bath solution lacking Mg^{2+} . Action potentials were fired by injecting into the presynaptic cell a 10-ms 1 nA current step at 0.1 Hz. The presynaptic cell's resting membrane potential was near -60 mV; the postsynaptic cell was voltage clamped at -60 mV. Arrows in the right-hand panels indicate spontaneous polysynaptic activity that was recorded simultaneously in both cells (the current trace, lower panel, goes off-scale). Such activity was much more frequent after tetanus. Calibration bars refer to both right- and left-hand panels. The dotted lines in the upper panels indicate 0 mV. *b*, Plot of peak e.p.s.c. amplitude as a function of time for the same experiment as in *a*. The arrow indicates the time at which the sequence of tetani was applied. Immediately before this the bath was changed to a 0 mM Mg^{2+} solution, and then immediately afterward back to the control 10 mM Mg^{2+} solution. The straight lines through the data points represent the mean amplitudes over the time windows indicated by the horizontal positions of the lines.

METHODS. Pyramidal neurons from fields CA1 and CA3 of the hippocampi of newborn Long Evans rat pups were maintained in dissociated cell culture as previously described²¹, except that 50 μ M D-APV (D-2-amino-5-phosphonopivalic acid) was added to the culture medium after about a week to block possible potentiation due to spontaneous firing. For electrophysiology, the bath solution comprised (in mM) NaCl (137), KCl (5), $CaCl_2$ (3), $MgCl_2$ (10 or 0), D-glucose (10), HEPES-NaOH buffer (5), pH 7.3, osmolarity adjusted to 310 mOsm with sorbitol, to which was added 1 μ M glycine, 1 μ M strychnine and 100 μ M picrotoxin. The tips of the patch electrodes were filled with a solution comprising (in mM) $KMeSO_4$ (150), KCl (5), EGTA (10), HEPES-KOH buffer (10), pH 7.2, and electrodes were back-filled with the same solution plus 100 μ g ml^{-1} nystatin, diluted from



a stock of 20 mg nystatin per ml of DMSO. Patch electrodes were Sylgarded and polished and had resistances of 1–3 M Ω with these solutions. After obtaining G Ω seals on a pair of cells, test pulses were applied until the access resistances stabilized (typically 30–50 min after forming the seals); access was checked for constancy throughout the experiment. Recordings were made with an Axopatch 1A, filter setting 1 kHz. The bath was perfused at about 1 $ml\ min^{-1}$ (bath volume 1 ml). Electrode offset potentials were within 2 mV at the end of the experiment. All measurements were made at room temperature (23–25 $^{\circ}C$).

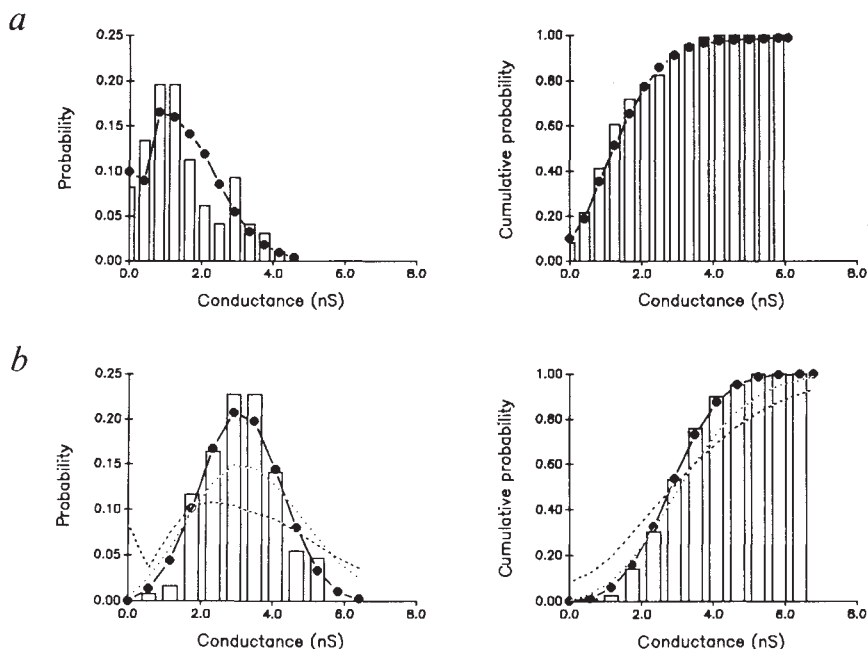


FIG. 2 Frequency histograms of e.p.s.c. conductance before and after inducing LTP are best fitted by assuming that LTP expression originates in an increase in release probability, p . *a*, Control (pre-tetanus) histograms. *b*, Post-tetanus histograms. (Same experiment as in Fig. 1.) In both *a* and *b* the right panel is the normalized cumulative histogram of the histogram in the left panel. In *a* the filled circles represent a binomial distribution fitted as described in the text. In *b* the filled circles were calculated from the distribution in *a* assuming that only p is scaled by the potentiation factor f , with no free parameters. The dotted line and the broken line were calculated assuming that the parameters N and a , respectively, are scaled by f .

artefact due to changing recording conditions. For example, the presynaptic action potential remained constant (Fig. 1*a*), and the bath was perfused at a continuous rate, preventing changes in the concentrations of possible modulatory substances²⁶.

Long-lasting potentiation was seen in 4 cell pairs out of 13 stable, connected pairs studied with the perforated patch technique. However, even in experiments in which the evoked e.p.s.c. was not potentiated, the above protocol usually caused a long-lasting increase in the general polysynaptic activity of cells in the culture. An example of this is visible in Fig. 1*a*, after tetanus (arrows). This is probably due to the potentiation of synapses other than those between the cell pair under study, and may arise in heterogeneity in the ability of synapses to support LTP²¹.

Our strategy was to carry out a quantal analysis of the pre-tetanus responses then to test, without estimating additional parameters, three alternative mechanisms for the potentiation by predicting the amplitude histogram obtained after potentiation had developed. The potentiated response in Fig. 1 could be due to $f = 2.1$ -fold scaling of N (the number of release sites), p (the quantal release probability) or a (the mean quantal size); as each of these factors enters the equation for the synaptic current size histograms in a different way, scaling each makes different predictions about the shape of the histograms. Thus we may determine which of the alternative theories best accounts for the data.

The probability of observing a synaptic current of a particular amplitude describing the amplitude histogram is given by the binomial distribution¹⁴ suitably modified to take account of the observed variability in the m.e.p.s.c. size²⁰. If a is taken to be the average value for the quantal size in culture (0.8 nS), the pre-tetanus binomial parameters N and p can be estimated from the standard equations for the mean $m = 1.62$ nS and variance $v = 1.073$ nS² of the amplitude histogram (Fig. 2*a*):

$$m = aNp$$

$$v = a^2 Np(1-p) + a^2 Npc_m^2.$$

The second term gives the contribution to the overall variance of fluctuations in the m.e.p.s.c. size; the coefficient of variation of m.e.p.s.c.s is c_m and has an average value of 0.5 in culture and 0.42 in slices²⁰. Solving these equations yields $N = 4.8$, $p = 0.42$. As N should be an integer, N was constrained to be 5 (rather than the calculated 4.8) and p was set to 0.37 (instead of 0.42).

The predicted and observed pre-tetanus histograms are shown superimposed in Fig. 2*a*; they are not significantly different at the 0.2 level (χ^2 goodness-of-fit test); that is, worse fits would occur by chance more than one time in five even if the histograms were generated by the same underlying mechanism. Although we could not in these experiments collect sufficient quantities of pre-tetanus data to carry out a very accurate quantal analysis, the fit exhibited in Fig. 2*a* confirms our previous conclusion²¹ that the Del Castillo and Katz formalism, with the m.e.p.s.c. variability included, accounts for the observed amplitude histogram.

Figure 2*b* shows the result of assuming that the mechanism of LTP is an f -fold increase in the release probability p , with N and a unchanged; the predicted histogram that results is superimposed on the observed post-tetanus amplitude histogram. The predicted and observed histograms are not significantly different at $P < 0.2$. The two other alternative mechanisms, an f -fold increase in N or in a make predictions that differ from the observed histograms at $P < 0.05$ and 0.01 levels of significance, respectively (dotted and broken lines in Fig. 2*b*).

A convenient way to compare predicted and observed histograms uses the coefficient of variation (and its squared value h)

$$c = \sqrt{h} = \sqrt{v/m}$$

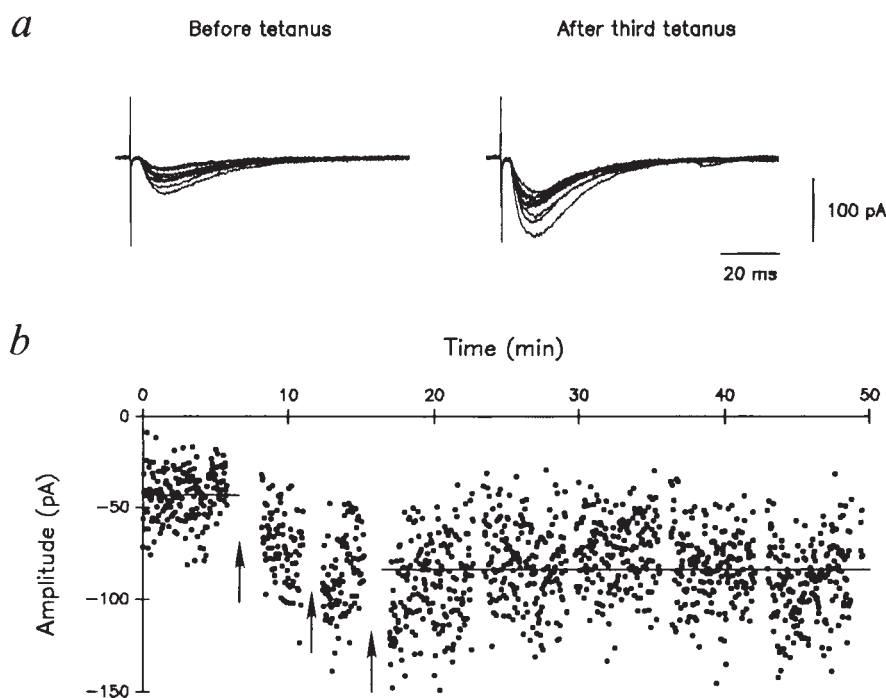
introduced in this context by Del Castillo and Katz¹⁴, and modified here to take account of m.e.p.s.c. variability:

$$h(N, p) = (1 + c_m^2 - p)/(Np). \quad (1)$$

Note that a cancels in the right-hand side of this equation, so

FIG. 3 Long-term potentiation of excitatory synaptic transmission in a patch-clamped hippocampal slice. *a*, Superimposed e.p.s.c.s, elicited by stimulation of the Schaffer collateral pathway at 0.5 Hz, recorded from a cell in area CA1 before (left) and after (right) tetanus-induced LTP. The neuron was whole-cell voltage-clamped at the soma at -70 mV. Calibration bars refer to both panels. *b*, Plot of the amplitudes of evoked e.p.s.c.s as a function of time during the experiment. At each arrow the patch amplifier was switched to current clamp mode and two 200-ms 100 Hz tetani were delivered 5 s apart. The amplifier was then returned to voltage clamp mode and testing resumed. The straight lines through the data points represent the mean amplitudes over the time windows indicated by the horizontal positions of the lines.

METHODS. Transverse slices (300–400 μ m) were cut from the hippocampi of 14–21-day-old Long Evans rats using a vibratome, submerged and continuously perfused at 1–2 ml min⁻¹ with oxygenated Krebs solution comprising (in mM) NaCl (120), KCl (3), MgCl₂ (1.2), CaCl₂ (2.5), NaHCO₃ (23), NaH₂PO₄ (1.2), D-glucose (11), to which 100 μ M picrotoxin was added. Patch electrodes contained (in mM) Cs-methylsulphate (120), CsCl (5), EGTA (0.5), MgCl₂ (2), Mg-ATP (2), sorbitol (10), HEPES-CsOH buffer (10), pH 7.3. Gigaohm seals were obtained with unpolished electrodes (resistances 3–5 M Ω) on the somas of cells in region CA1 and recordings made in the whole-cell with an Axopatch 1A, filter setting 2 kHz. A monopolar stimulating electrode (35 μ m diameter teflon-coated Pt-Ir wire) was positioned 200–300 μ m away in stratum radiatum and the stimulus intensity set just above the level that



gave mostly failures of synaptic transmission; this setting was subsequently left unchanged throughout the experiment. The stimulus duration for low frequency stimulation was 100 μ s; for tetani it was 200 μ s. Electrode offset potentials were within 2 mV at the end of the experiment. Experiments were performed at room temperature (23–25 $^{\circ}$ C).

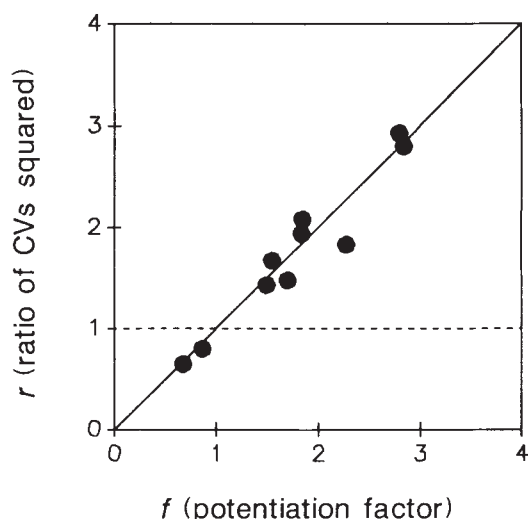


FIG. 4 A plot of r against f (defined in the text) for 10 different slice LTP experiments is best fitted by a model that locates the expression of LTP in the presynaptic terminal. The diagonal line is the prediction of a Poisson scheme in which LTP is expressed presynaptically, the horizontal line the prediction for postsynaptic expression. The two points for which f is less than unity were obtained from experiments in which 100 μ M D-APV was present during the tetanus, and a long-lasting depression resulted.

that the value of h is not altered by a potentiation mechanism that operates through increases in a . According to this equation, if a increased f -fold, $h(N, p) = 0.48$ (the pre-tetanus value), whereas $h(fN, p) = 0.23$, and $h(N, fp) = 0.12$. The observed value for h in this experiment was 0.08. This supports the above conclusion that LTP in culture involves an increase in p rather than in N or a . The same conclusion was reached in full quantal analyses of three other cell pairs in culture that displayed long-lasting potentiation.

Because the potentiation just described is enduring, and requires activation of synapses under conditions that permit calcium influx through the NMDA (*N*-methyl-D-aspartate) receptor channels, it is like LTP^{7,8}. LTP is, however, normally studied in relatively intact hippocampal circuits, not dissociated cultures, so one cannot be sure that the mechanism discussed above operates in the same way as actual LTP. For this reason, we experimented with whole-cell clamped neurons in hippocampal slices, where the properties of LTP are well known.

Experiments like those just described, in which one records with the perforated patch technique from defined cell pairs, are difficult in the slice as one can only rarely find monosynaptic connections¹⁶. Accordingly, we used the conventional whole cell ruptured patch technique to record from single postsynaptic pyramidal neurons while repetitively exciting axons of the Schaffer collateral pathway with a small extracellular stimulating electrode^{27,28}. An example is shown in Fig. 3. Panel *a* shows superimposed e.p.s.c.s recorded while voltage clamping the neuron at -70 mV, measured before and after tetanus. Panel *b* plots the e.p.s.c. amplitude as a function of time during the experiment. Tetani were delivered at each of the arrows in Fig. 3*b* while maintaining the cell in current clamp. This protocol induced a long-lasting potentiation of the e.p.s.c.; the mean control amplitude was 43.3 pA, and the mean following three series of tetani was 83.8 pA.

In the above experiment the stimulus level was set just above the threshold for synaptic transmission. We knew, however, that multiple axonal inputs onto the postsynaptic cell were being stimulated. This situation complicates the quantal analysis¹⁵.

Analysis in hippocampal slices

We have taken two approaches to extending our analysis to the hippocampal slice. First, we have used a procedure, described

below, that takes account of the possible nonuniformity of the stimulated inputs to a neuron. Second, we have simply applied the standard equations of quantal analysis, and have relied upon goodness-of-fit tests to determine if the application of these equations is valid. These approaches are described in turn.

The first approach is best described by presenting a graph that plots a quantity r as a function of the factor f by which the mean e.p.s.c. size is potentiated. The quantity r , like f , is determined directly from the experimental data and is defined as the ratio

$$r = h/h_p$$

where h (defined earlier in equation (1)) is the squared value of the coefficient of variation before a tetanus and h_p is the corresponding quantity in the potentiated slice. In the limiting case where p is small (the Poisson limit), r is, according to equation (1), equal to f if the mechanism of LTP is an f -fold change in either N or p (presynaptic mechanism) and equal to 1 if LTP results from a scaling of a (postsynaptic mechanism). Figure 4 shows such an r/f plot for 10 slice experiments and unambiguously accords with the predictions of the presynaptic mechanism. Note that in two cases f was less than unity; in these experiments the tetanus was applied with 100 μ M D-APV in the bath, which resulted in a long-lasting depression.

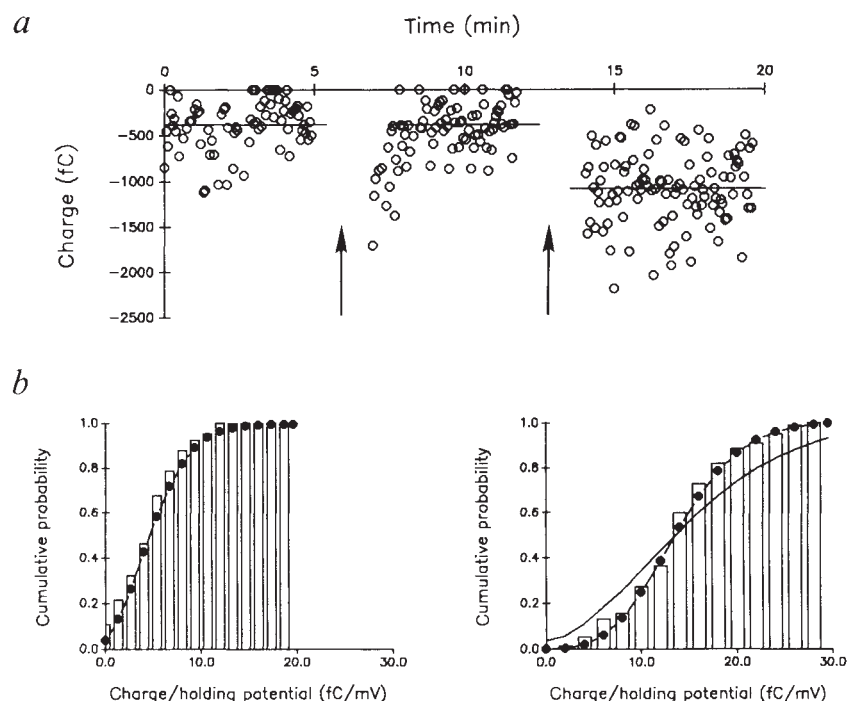
The r/f plot can be generalized to many situations with multiple inputs to a neuron in which a standard quantal analysis is not applicable. It can be shown in these cases that points will fall on or to the left of the diagonal ($r = f$) if the LTP mechanism is presynaptic (an increase in either N or p), and will fall on a horizontal line ($r = 1$), or between that line and the diagonal, if the mechanism is postsynaptic (an increase in a). However, analysis reveals a special case for which the r/f plot does not correctly distinguish between a pre- and postsynaptic mechanism for LTP. Suppose that individual postsynaptic membranes could exist in only two states, 'on' and 'off'; for an 'off' synapse release of a quantum of neurotransmitter would result in essentially no response, and for an 'on' synapse, each quantum would give a standard-sized response. The mechanism of LTP in this model would be the postsynaptic switching (for example, by increasing the sensitivity of receptors in the postsynaptic membrane) from 'off' to 'on'. Such a switch would appear in our analysis to be an increase in N , and thus to be presynaptic, whereas in fact the site of LTP expression would be postsynaptic. From our slice experiments alone we would not be able to distinguish between the presynaptic mechanism and this particular postsynaptic mechanism. The observations in culture, however, are not consistent with the postsynaptic mechanism because LTP there resulted from an increase in the release probability p , not N .

We conclude that our observations in slices, like those in culture, are consistent with the expression of LTP being primarily presynaptic. Because points fall along the diagonal, indicating that release is described by the Poisson limit, we cannot here, as we could in some of the experiments on neuron pairs in culture, distinguish between effects on N and p .

An alternative way to test our conclusion is to carry out a full quantal analysis of transmission in these slice experiments: if the mechanism is indeed presynaptic and the quantal parameters sufficiently constant to give data points that fall along the diagonal in the r/f plots, then the distribution of synaptic current sizes should be well described by the standard equations of quantal analysis, and one should be able to predict the distribution of potentiated synaptic current sizes with the quantal parameters obtained from the unpotentiated distribution. If the conclusion is incorrect, then the observed distribution of synaptic current sizes should be a sum of Poisson distributions, and may not, in general, be adequately described by a single Poisson distribution.

Figure 5*a* shows a slice experiment in which the postsynaptic response for a small stimulus is plotted as a function of time

FIG. 5 Histograms of e.p.s.c. size in the slice are best fitted by a Poisson model in which LTP expression originates in a (presynaptic) increase in the mean quantal content, Np . *a*, Plot of the charge carried by evoked e.p.s.c.s as a function of time during the experiment. Charge was calculated by integrating individual e.p.s.c.s over an 80–90-ms time window starting at the foot of the e.p.s.c. Test stimuli were delivered at 0.5 Hz while voltage clamping the soma at -70 mV. At the first arrow, three trains of 200-ms 100 Hz tetani were delivered at 5-s intervals while still voltage clamping the soma at -70 mV. At the second arrow the same tetanus protocol was applied while the voltage clamping at -20 mV. The straight lines through the data points represent the mean amplitudes over the time windows indicated by the horizontal positions of the lines. *b*, Cumulative probability histograms derived from the data in *a* before (left) and after (right) inducing LTP. Note that the charge carried by the e.p.s.c.s has been normalized by dividing by the holding potential (-70 mV). The filled circles on the left are a fit of the Poisson distribution to the experimental histogram. The filled circles on the right were calculated from the Poisson distribution assuming that LTP scales only Np by f , with no free parameters. The solid curve on the right is the prediction of a Poisson distribution assuming that only parameter a is scaled by f .



during the experiment. Note that charge, rather than amplitude, is measured because the former is a quantity less affected by cable attenuation in the dendrites^{17,29}. At the point indicated by the first arrow, the inputs to the cell were stimulated repetitively but the postsynaptic neuron was maintained by the voltage clamp at -70 mV; the response did not potentiate after this procedure, as expected for LTP^{3–6}. At the second arrow, the cell's inputs were again stimulated while the cell was held at -20 mV, a value that should maximize calcium influx through the NMDA receptor channels and produce LTP^{7,8}. The mean response after this tetanus was potentiated by a factor $f = 2.84$. This particular experiment was chosen for illustration of the full quantal analysis because failures were apparent in the control records, which provides an additional check on the validity of the fitted parameters.

Figure 5*b* shows the histogram of observed synaptic responses for this experiment with superimposed predictions of the Poisson distribution, using appropriate values for a and Np estimated from the mean and variance of the raw data, and again taking into account the measured variance in m.e.p.s.c. size²⁰. A χ^2 goodness-of-fit test detects no difference between these distributions at the 0.2 significance level. The quantal size estimated by the ratio variance/mean for this experiment is 2.1 fC mV⁻¹ before (and after) the tetanus. For six experiments, the average quantal size was 1.35 ± 0.37 fC mV⁻¹ before the tetanus and 1.41 ± 0.28 fC mV⁻¹ in the potentiated slice (mean $\pm$ s.d.). These estimates should be compared with the mean value for a directly determined in slices of 1.5 fC mV⁻¹ with a standard deviation of 0.5 fC mV⁻¹ (ref. 20). The value for a estimated from the histogram is thus within the range of quantal sizes determined independently.

If the pre-tetanus value for Np is multiplied by f , the predicted histogram, superimposed on the experimental one after potentiation, is shown in Fig. 5*c*. Again the χ^2 goodness-of-fit test detects no difference between the observed and predicted histograms at the 0.2 significance level. If a , rather than Np , is multiplied by f , the predicted and observed histograms differ at the 0.01 significance level, so the differences are statistically different (Fig. 5*c*, smooth curve). Similar results were found in four other slice experiments in which we carried out full quantal analyses. We conclude that the potentiation mechanism is presynaptic, and that either N or p increases following the tetanus.

The longest LTP observed in our slice experiments is 98 min, and our conclusions hold up to that time. We cannot, however, exclude the possibility that slower developing phases of LTP use some different mechanism. Because our experiments were done at room temperature (23 – 25 °C), and we do not know the temperature coefficients of the various possible phases of LTP, we cannot even place limits of the validity of our conclusions for the hippocampus at body temperature. From just after the tetanus up to 1 h after, only a single phase of LTP was apparent in our experiments, and the statistical properties of the postsynaptic response were essentially invariant.

Discussion

Our conclusion is that the mechanism of LTP is presynaptic rather than postsynaptic, and probably involves p instead of N . We have not, however, attempted to place specific limits on the extent to which intermediate cases, in which two or three of the quantal parameters are affected by LTP, might hold. Specifically, we cannot exclude the possibility that some small fraction of LTP is postsynaptic, or that both N and p increase with the p effect predominating. Because the potentiation factor f enters squared if the LTP mechanism is postsynaptic and linearly if it is presynaptic, a large contribution from the postsynaptic site can be excluded. N as well as p could be affected significantly, and we would not be able to detect this.

Our conclusion that the expression of LTP is, under the conditions of our experiment, mainly presynaptic, directs attention to mechanisms whereby retrograde signals from the postsynaptic cell cause enhanced presynaptic release of neurotransmitter^{13,30,31}.

Note added in proof: Our observations and conclusions are in accord with those reported by R. Malinow and R. W. Tsien³². □

Received May 15; accepted 25 July 1990.

1. Bliss, T. V. P. & Lomo, T. *J. Physiol., Lond.* **232**, 331–356 (1973).
2. Bliss, T. V. P. & Lynch, M. in *Long-Term Potentiation: From Biophysics to Behavior* (eds Landfield, P. W. & Deadwyler, S. A.) 3–72 (Liss, New York, 1988).
3. Wigstrom, H., Gustafsson, B., Huang, Y.-Y. & Abraham, W. C. *Acta physiol. scand.* **126**, 317–319 (1986).
4. Kelso, S. R., Ganong, A. H. & Brown, T. H. *Proc. natn. Acad. Sci. U.S.A.* **83**, 5326–5330 (1986).
5. Malinow, R. & Miller, J. P. *Nature* **320**, 529–530 (1986).
6. Sastry, B. R., Goh, J. W. & Auyeung, A. *Science* **232**, 988–990 (1986).

7. Collingridge, G. L. & Bliss, T. V. P. *Trends Neurosci.* **10**, 288–293 (1987).
8. Gustafsson, B. & Wigstrom, H. *Trends Neurosci.* **11**, 156–162 (1988).
9. Skrede, K. K. & Maltre-Sorensen, D. *Brain Res.* **208**, 436–441 (1981).
10. Dolphin, A. C., Errington, M. L. & Bliss, T. V. P. *Nature* **297**, 496–498 (1982).
11. Muller, D., Joly, M. & Lynch, G. *Science* **242**, 1694–1697 (1988).
12. Kauer, J. A., Malenka, R. C. & Nicoll, R. A. *Neuron* **1**, 911–917 (1988).
13. Davies, S. N., Lester, R. A. J., Reymann, K. G. & Collingridge, G. L. *Nature* **338**, 500–503 (1989).
14. Del Castillo, J. & Katz, B. *J. Physiol., Lond.* **124**, 560–573 (1954).
15. Redman, S. J. *Physiol. Rev.* **70**, 165–198 (1990).
16. Sayer, R. J., Redman, S. J. & Andersen, P. J. *Neurosci.* **9**, 840–850 (1989).
17. Rall, W. & Segev, I. in *Voltage and Patch Clamping with Microelectrodes* (eds T. G. Smith et al.) 191–215 (American Physiological Society, Bethesda, 1985).
18. Brown, T. H., Wong, R. K. S. & Prince, D. A. *Brain Res.* **174**, 194–199 (1979).
19. Yamamoto, C. *Expl Brain Res.* **46**, 170–176 (1982).
20. Bekkers, J. M., Richerson, G. B. & Stevens, C. F. *Proc. natn. Acad. Sci. U.S.A.* **87**, 5359–5362 (1990).
21. Bekkers, J. M. & Stevens, C. F. *Nature* **341**, 230–233 (1989).
22. Voronin, L. L. in *Synaptic Plasticity in the Hippocampus* (eds Haas, H. L. & Buzsaki, G.) 27–30 (Springer, Berlin, 1988).
23. Foster, T. C. & McNaughton, B. L. *Soc. Neurosci. Abstr.* **15**, 775 (1989).
24. Horn, R. & Marty, A. *J. gen. Physiol.* **92**, 145–159 (1988).
25. Hamill, O. P. et al. *Pflugers Arch.* **391**, 85–100 (1981).
26. Forsythe, I. D., Westbrook, G. L. & Mayer, M. L. *J. Neurosci.* **8**, 3733–3741 (1988).
27. Edwards, F. A., Konnerth, A., Sakmann, B. & Takahashi, T. *Pflugers Arch.* **414**, 600–612 (1989).
28. Blanton, M. G., Lo Turco, J. J. & Kriegstein, A. R. *J. Neurosci. Meth.* **30**, 203–210 (1989).
29. Bekkers, J. M. & Stevens, C. F. *Progr. Brain Res.* **83**, 37–45 (Elsevier, Amsterdam, 1990).
30. Piomelli, D. et al. *Nature* **328**, 38–43 (1987).
31. Williams, J. H., Errington, M. L., Lynch, M. A. & Bliss, T. V. P. *Nature* **341**, 739–742 (1989).
32. Manilow, R. & Tsien, R. W. *Nature* **346**, 177–180 (1990).

ACKNOWLEDGEMENTS. We thank Drs L. Chavez-Noriega and G. Richerson for their suggestions on the slice technique. This work was supported by the Howard Hughes Medical Institute and the NIH (C.F.S.).

LETTERS TO NATURE

Detection of diffuse interstellar bands in the infrared

C. Joblin*, J. P. Maillard†, L. d'Hendecourt* & A. Léger*

*Groupe de Physique des Solides, Université Paris 7, CNRS, 2 place Jussieu, 75251 Paris Cedex 05, France

†Institut d'Astrophysique de Paris, CNRS, 98 bis, Boulevard Arago, 75014-Paris, France

THE diffuse interstellar bands (DIBs) are spectral absorption features caused by interstellar matter but not identified with any known atomic or molecular lines. Since 1975, when 39 such bands were known¹, many more have been detected, and in 1988, about 80 were listed². They range in frequency from visible (4,430 Å) to very near infrared wavelengths (8,650 Å) (ref. 3), but their carriers—the atoms or molecules that cause them—remain unknown. We have found two new DIBs in the near infrared, near 11,800 Å and 13,200 Å, in the spectra of reddened O, B and A stars. Their full widths at half maximum are 2.7 ± 0.3 Å and 4.0 ± 0.5 Å respectively, and their equivalent widths are 60 and 135 mÅ per magnitude of extinction. The wavelength of the second DIB is a factor of 1.5 longer than that of the longest wavelength band previously identified, which may be an important constraint on the nature of the carriers of these two bands.

Among the candidates for the carriers of the DIBs, carbonaceous molecules are particularly attractive because carbon is abundant and its bonds can survive the ultraviolet radiation in the diffuse interstellar medium. Several candidates have been proposed: carbon chains⁴, polycyclic aromatic hydrocarbons (PAHs)^{5,6}, which are thought to be abundant in the interstellar medium, and fullerene ions^{7,8}. Kroto⁹ has shown theoretically that only a few such ions can form stable clusters, mainly C_{50}^+ , C_{70}^+ and especially the most symmetric one C_{60}^+ , a result in agreement with laboratory experiments⁷. Such species could be abundant in the interstellar medium and may be responsible for some of the DIBs. No spectroscopic data on C_{60}^+ are available but according to calculations using the Hückel model⁸, C_{60}^+ has a ground-state transition at 4,500 Å ($\pm 20\%$) that is weakly allowed and could give rise to the observed DIB at 4,430 Å. The same calculations imply that there should also be a very

strong DIB in the infrared⁸ owing to an allowed transition at 1.2 μm ($\pm 20\%$). This led us to look for interstellar features in the near infrared.

Our observations were carried out on the 3.6-m Canada–France–Hawaii telescope at Mauna Kea using a Fourier-transform infrared spectrometer with a resolution of $\sim 10^4$ (resolution element 0.9 cm^{-1}). From Herbig¹, we selected reddened stars exhibiting DIBs in the visible. The stars studied are listed in Table 1—they have different spectral types as well as various values of colour excess, $E(B-V)$. During the first run (September 1988), we studied two objects using the J, H and K filters and extracted two possible interstellar features in the J filter but none in the others. In the second run (November 1989), we observed four objects using only the J filter.

To remove most of the telluric absorptions (mainly H_2O absorption bands), we divided the spectrum of an object by the spectrum of a reference star. A good reference should be located at the same air mass as the object and also be of the same spectral type. Moreover, to allow for correct removal of stellar features without cancelling truly interstellar ones, the standard star should show no reddening. We tried to fulfil these criteria as much as possible. The remaining features in the ratio spectra are of three kinds: (1) incompletely removed telluric lines; (2) intrinsic stellar features that are not strictly compensated (for example, Fig. 1A shows He I features visible in the ratio of the spectra of object 6 and a reference of spectral type B1 III). Both these kinds of features can be identified because they are prominent in the raw spectrum of objects and references; (3) interstellar features. We consider a feature to be interstellar if it is present in the most reddened stars (applying a 3σ detection criterion) and absent in references, if its strength varies with $E(B-V)$, and if its position from one line of sight to another can be explained by a reasonable Doppler shift. Typical interstellar clouds have a velocity dispersion of $< 50\text{ km s}^{-1}$, which gives a maximum variation from one line of sight to another of 2 Å for a mean wavelength of 12,000 Å.

We found two such features. The stronger band lies between 13,174 Å ($7,588.5\text{ cm}^{-1}$) and 13,176 Å ($7,587.6\text{ cm}^{-1}$) (wavelengths λ are measured in air at standard temperature and pressure and wavenumbers σ are measured in vacuum: $\lambda(\text{Å}) = \sigma(\text{cm}^{-1}) \times 10^8/n$ where n is the index of refraction of air, 1.000273) and has a full width at half maximum (FWHM) of 4.0 ± 0.5 Å. The wavelength range of the weaker band is 11,797 Å

TABLE 1 Stars observed

Object	Name	Spectral type	Visual magnitude	J-filter magnitude	$E(B-V)$
1	BD +40°4220	O7f	9.20	4.67	2.00
2	HD183143	B7Ia	6.87	3.56	1.28
3	HD194279	B1.5a	7.02	4.23	1.22
4	HD20041	A0Ia	5.92	4.06	0.73
5	HD223385	A3Ia	5.42	3.50	0.59
6	HD24398	B1Ib	2.85	2.65	0.34

($8,474.3\text{ cm}^{-1}$)– $11,798\text{ Å}$ ($8,473.9\text{ cm}^{-1}$) and the FWHM is $2.7 \pm 0.3\text{ Å}$. We have not tried to correct these wavelengths for the velocities of the interstellar clouds in front of the stars.

Curves *b* and *c* in Fig. 1*B*, *C* show the two features in the ratio spectra of objects 1 and 2, respectively. Figure 1*Ba*, *Ca* show the ratios of references with other references giving an estimate of the remaining spurious features and Fig. 1*Bd*, *Cd* are the raw spectra of the references showing no clear structure at these wavelengths in unreddened stars. The shape of the two features seems slightly asymmetric owing to possible contributions from several clouds along the line of sight or to overlapping telluric bands that are not completely removed. The latter also

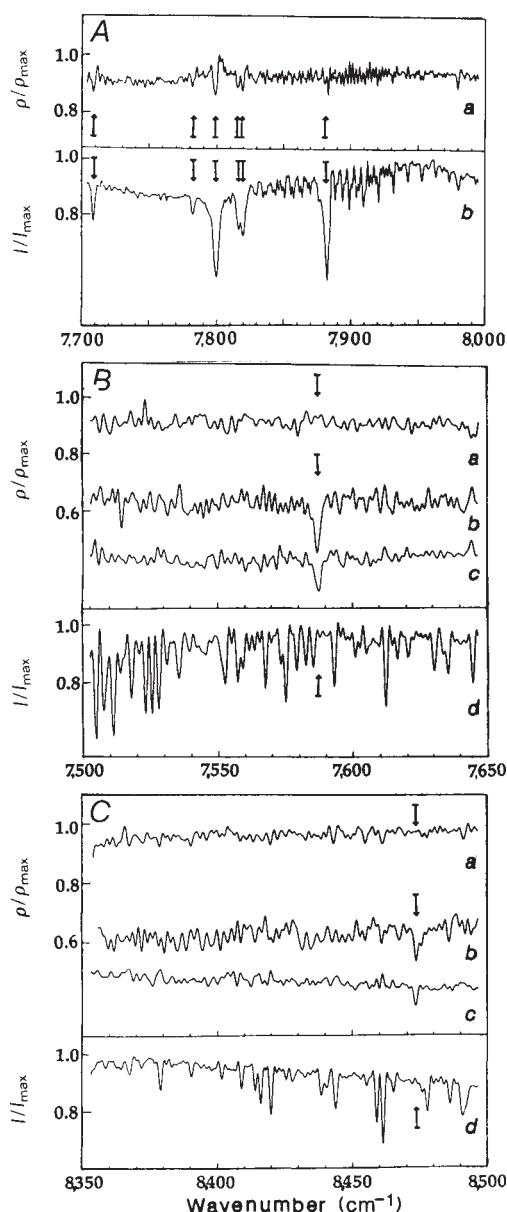


FIG. 1 Spectra with *A*, stellar features and *B*, diffuse interstellar features. *Aa* shows the incomplete removal of the He I lines when the spectrum from object HD24398 is divided by the spectrum of the reference star HD23180 (spectral type B1III), shown in *Ab*. *Ba* and *Ca* show the ratios of the spectra from some reference stars (HD33111 (spectral type A3III)/HD216627 (spectral type A3V) and HR1203 (B1Ib)/HR1122 (BSIII)). *Bb* and *Bc* show the interstellar band at 1.32 μm for BD +40°4220 and HD183143. *Cb* and *Cc* show the interstellar band at 1.18 μm for BD +40°4220 and HD183143. *Bd* and *Cd* show that the DIBs at 1.32 μm and 1.18 μm are absent in the spectrum of reference star HD23180 and of the coadded references HR7924 (spectral type A1ae) and HR1122, respectively. Each spectrum is normalized at its maximum value. For clarity, the curves have been shifted vertically.

introduces an error in the evaluation of the equivalent widths. To estimate this error, we have compared the ratios obtained with different references and the calculated uncertainties take this into account, as well as the uncertainty owing to noise.

In Fig. 2, the two equivalent widths are plotted against $E(B-V)$. The correlation coefficients are 0.91 and 0.71 and the equivalent widths are 135 and 60 mÅ per magnitude of visual extinction. The features for object 4 seem to be as strong as for objects towards which the extinction is twice as large. Herbig also observed this for the DIB at $5,780\text{ Å}$.

In addition to a correlation with reddening, we checked that these features are probably not of stellar origin using a simple model of stellar atmosphere. For any given equivalent width and wavenumber, we obtain a condition of minimum abundance on the absorbing element assuming an oscillator strength ≤ 1 , where we consider an element to be an atom with a certain degree of ionization and in a certain energy level. We searched tabulations of atomic and ionic transitions¹⁰ but did not find any suitable candidate in a range of 5 cm^{-1} around the observed positions.

The two observed features are most likely not atomic or ionic transitions from simple atoms in the interstellar medium because all tabulated transitions around these two frequencies ($\pm 5\text{ cm}^{-1}$) involve high energy levels, which cannot be sufficiently populated by the interstellar radiation field or by collisions.

What about molecular carriers? The observed transitions are certainly electronic because allowed vibrational transitions have low oscillator strengths and the position of the bands requires that they would arise from combination modes or overtones, which have a much weaker intensity. In the interstellar medium, neutral molecules are mostly in their ground state. For molecules fulfilling Platt's criterion¹¹ (such as those having $\Pi \rightarrow \Pi^*$ transition), the wavelength of the first allowed electronic transition from the ground state is proportional to the molecular size; such a transition in the infrared implies quite large molecules. This model¹¹ gives a linear size of 28 Å (25 Å) for a band at $7,588\text{ cm}^{-1}$ ($8,474\text{ cm}^{-1}$). For pericondensed (the most compact arrangements of aromatic rings) PAHs, which are expected to be abundant interstellar carbonaceous molecules and candidates for the carriers of the DIBs^{5,6}, such a dimension implies nearly 350 carbons per molecule¹², compatible with the size distribution of these molecules derived from infrared data¹³. But this result does not apply to radicals, which can have transitions at lower energies. Given that interstellar species can be partially ionized and dehydrogenated, and therefore can be radicals, the two transitions could arise from smaller species.

As mentioned above, calculations using the Hückel model imply that C_{60}^+ , if responsible for the DIB at $4,430\text{ Å}$, could provide a very strong absorption feature at 1.2 μm ($\pm 20\%$). The two detected bands are much weaker than expected (and the observed ratio of the oscillator strengths at 1.32 μm and 0.4430 μm is 2×10^{-2} whereas the expected ratio between an allowed and a forbidden transition should be much greater than unity). Although it is tempting to conclude that C_{60}^+ cannot

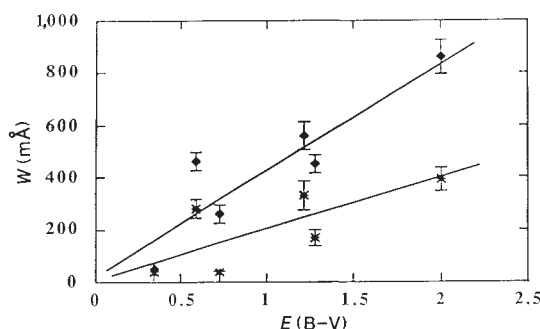


FIG. 2 Equivalent width of the 1.32-μm feature (diamonds) and the 1.18-μm feature (asterisks) plotted against selective extinction $E(B-V)$.

account for the 4,430-Å band, the spectral range observed (the J band: 1.2 μm (±12%)) is not wide enough to cover fully the region in which the C₆₀⁺ band might be expected owing to uncertainty on the position from calculations using the Hückel model. Experimental data are needed to try and constrain the abundance of C₆₀⁺ in the interstellar medium.

Further studies are necessary to address, for example, whether or not there is a relationship between these two bands and the families of DIBs recently determined in the visible¹⁴. It is also important to determine whether these bands are present in molecular or diffuse clouds and if they pertain to neutral or ionized species. □

Received 27 February; accepted 27 June 1990.

- Herbig, G. H. *Astrophys. J.* **196**, 129–160 (1975).
- Herbig, G. H. *Astrophys. J.* **331**, 999–1003 (1988).
- Sanner, F., Snell, R. Van den Bout, P. *Astrophys. J.* **226**, 460–468 (1978).
- Douglas, A. E. *Nature* **269**, 130–132 (1977).
- Van der Zwet, G. P. & Allamandola, L. J. *Astr. Astrophys.* **146**, 77–80 (1985).
- Léger, A. & d'Hendecourt, L. *Astr. Astrophys.* **146**, 81–85 (1985).
- Kroto, H. W. in *Polycyclic Aromatic Hydrocarbons and Astrophysics* (eds Léger, A., d'Hendecourt, L. & Boccara, N.) 197–206 (Reidel, Dordrecht, 1987).
- Léger, A., d'Hendecourt, L., Verstraete, L. & Schmidt, W. *Astr. Astrophys.* **203**, 145–148 (1988).
- Van der Zwet, G. P. & Allamandola, L. J. *Nature* **329**, 529–531 (1987).
- Kurucz, R. L. *Semi-empirical Calculation of gf Values* (Smithsonian Inst., Cambridge, 1981).
- Platt, J. R. *Astrophys. J.* **123**, 486–490 (1956).
- Puget, J. L., Léger, A. & Boulanger, F. *Astr. Astrophys.* **142**, L19–L22 (1985).
- Puget, J. L. & Léger, A. *Rev. Astr. Astrophys.* **37**, 161–198 (1989).
- Kreowski, J. & Walker, G. A. H. *Astrophys. J.* **312**, 860–867 (1987).

Extending the power of powder diffraction for structure determination

W. I. F. David

Neutron Science Division, ISIS, Rutherford Appleton Laboratory, Chilton, Oxon, OX11 0QX, UK

IN the modern techniques of high-resolution neutron and X-ray powder diffraction, an impressive amount of structural information may be obtained, allowing the refinement, using the Rietveld technique¹, of structures containing more than 100 positional parameters. Despite these technical advances and the renewed popularity of the technique, powder diffraction suffers from the inevitable loss of information imposed by the collapse of three dimensions of diffraction data onto the one dimension of a powder pattern. This can seriously frustrate the determination of unknown crystal structures. As a consequence of accidental or exact reflection overlap, only relatively few reflections may have uniquely determined intensities. (For example, the cubic reflections 550, 710 and 543 exactly overlap.) Here I present a maximum-entropy algorithm (previously discussed theoretically²) that evaluates, from first principles, the intensities of overlapping reflections in powder diffraction profiles. Indeed, a full three-dimensional reconstruction of the image may be retrieved when only ~20% of the reflections are uniquely determined. This precludes systems where the Patterson group is not the holosymmetric group (for example, the 4/*m* symmetry where *hkl* and *khl* reflections are overlapping but not equivalent.) Nevertheless, for most crystal structures with unit cell volumes of less than 1,000 Å³, this method should be useful.

Crystallography is rooted in the mathematics of the Fourier transform. The structure factor $F(\mathbf{h})$ (in general a complex quantity) that is associated with reflection $\mathbf{h}$ is the Fourier coefficient of the atomic structure $\rho(\mathbf{x})$

$$F(\mathbf{h}) = \int_V \rho(\mathbf{x}) \exp(2\pi i \mathbf{h} \cdot \mathbf{x}) d^3x$$

In a crystallography experiment the measured quantity is

$|F(\mathbf{h})|^2 = F(\mathbf{h})F^*(\mathbf{h})$ and the phase information is lost. $|F(\mathbf{h})|^2$ is the Fourier component of the Patterson function $P(\mathbf{u})$, the autocorrelation function of the scattering density (the interatomic vector map of the crystal structure)

$$|F(\mathbf{h})|^2 = \frac{1}{2} \int_V P(\mathbf{u}) \cos(2\pi \mathbf{h} \cdot \mathbf{u}) d^3u$$

In a powder diffraction experiment many reflections may overlap and, in general, measured observations are of the form

$$G_i = \sum_{n=1}^{N_i} J(\mathbf{h}_n) |F(\mathbf{h}_n)|^2 = \int_V P(\mathbf{u}) \left[\sum_{n=1}^N \left\{ \sum_{j=1}^{J(\mathbf{h}_n)} \cos(2\pi \mathbf{h}_n \cdot \mathbf{u}) \right\} \right] d^3u \quad (1)$$

where J is the number of symmetry equivalent reflections. Although the observations are no longer Fourier coefficients of any meaningful function, equation (1) is nevertheless linear and, if the Patterson map is divided into N_{pix} pixels may be expressed in the form

$$G_i = \sum_{k=1}^{N_{\text{pix}}} R_{ik} P_k \quad (i=1, \dots, N_{\text{obs}}) \quad (2)$$

The coefficients R_{ik} correspond to the quantity inside the square brackets of equation (1). The method proposed here is to maximize the 'Patterson entropy'

$$S = - \sum_{k=1}^{N_{\text{pix}}} P_k \ln(P_k/Q_k) \quad (3)$$

where Q_k is the *a priori* estimate (assumed flat) of the Patterson map, while maintaining agreement with the observed data D_i (standard deviation σ_i) through the χ^2 constraint

$$\sum_{i=1}^{N_{\text{obs}}} \frac{1}{\sigma_i^2} (D_i - G_i)^2 = N_{\text{obs}}$$

The maximum-entropy technique is a powerful method of image reconstruction that has been successfully applied to a number of different areas of research ranging from radio astronomy³ to crystallography. The application of maximum entropy to crystallography has been discussed by several authors (see, for example, refs 4 and 5). The above formulation of the maximum-entropy algorithm implicitly assumes that the scattering density is everywhere positive. Although this is true for X-ray diffraction analysis, negative Patterson densities may result from the analysis of neutron diffraction data. In such cases the Patterson map is divided into two arrays with intrinsically positive and negative contributions. Equations (2) and (3) are then recast as

$$G_i = \sum_{k=1}^{N_{\text{pix}}} R_{ik} (P_k^+ - P_k^-) \quad (i=1, \dots, N_{\text{obs}}) \quad (2a)$$

$$S = - \sum_{k=1}^{N_{\text{pix}}} [P_k^+ \ln(P_k^+/Q_k^+) + P_k^- \ln(P_k^-/Q_k^-)] \quad (3a)$$

The final step in the full three-dimensional reconstruction of the diffraction data is achieved by calculating individual Fourier components of the maximum-entropy Patterson function and thus separating the intensities of overlapping reflections

$$|F(\mathbf{h}_n)|^2 = \frac{1}{2} \int_V P_{\text{ME}}(\mathbf{u}) \cos(2\pi \mathbf{h}_n \cdot \mathbf{u}) d^3u \quad (4)$$

As an example of the method, I present an analysis of medium-resolution neutron powder diffraction data. Rutile (TiO₂) was selected because its relatively high tetragonal symmetry ($P4_2/mnm$: $a = 4.5929$ Å, $c = 2.9581$ Å) and the medium resolution of the data set collected ensures a reasonable degree of reflection overlap and because the presence of positive (O) and negative (Ti) scattering centres tests the generality of the method.

Diffraction data from a 5-cm³ sample of rutile were obtained

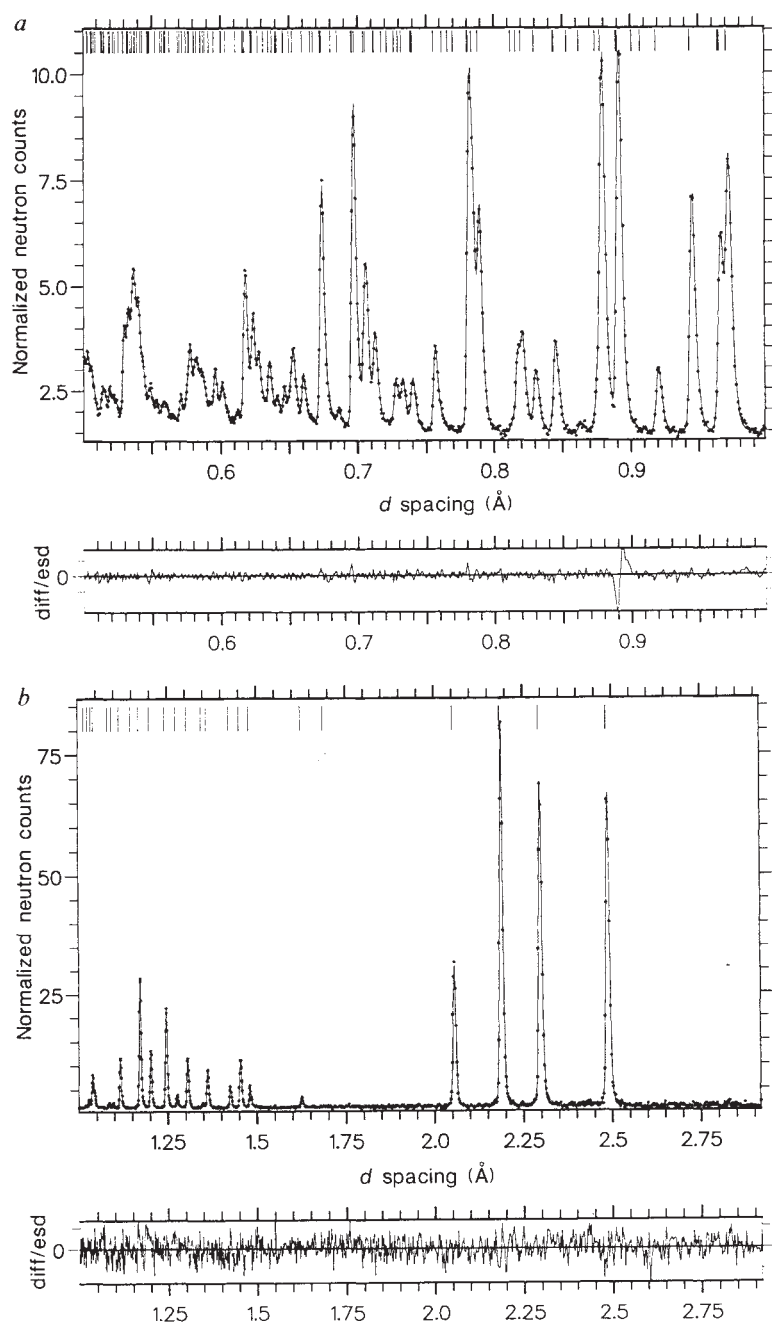


FIG. 1 The observed (dots), calculated (line) and difference profiles for TiO_2 . The d spacings range from *a*, 0.5 to 1 $\text{\AA}$ and *b*, 1 to 2.9 $\text{\AA}$. The positions of Bragg reflections are marked at the top and clearly indicate the severe peak overlap particularly at short d spacings; by eye, there are only 65 peaks in the full pattern. The difference profile is plotted as (observed – calculated profiles)/(standard deviation).

from the Polaris⁶ diffractometer at ISIS in 2 h. After normalization, the profile was fitted by least squares refinement using the technique devised by Pawley⁷; in addition to the Bragg peak position (determined by the size of the unit cell) and the peak width variables, the intensities of individual Bragg reflections were refined as separate variables with the proviso that reflections that were close together (typically <10% of the full width

at half maximum) were constrained to have equal structure-factor magnitudes. (This is the correct *a priori* Bayesian assumption.) 177 reflections were recorded in 130 clumps; 97 reflections were not overlapping. Comparison of the observed and calculated diffraction patterns (Fig. 1) indicates an excellent agreement ($\chi^2 = 1.371$). Figures 2*a–c* show sections of the Patterson map obtained by direct Fourier transformation of the equipartitioned squares of the structure-factor magnitudes obtained from the Pawley refinement. Broad features dominate the maps with clear evidence of oscillations resulting from Fourier truncation errors and splitting of interatomic vector peaks caused by misapportioning of Bragg intensities. It is not possible to locate the interatomic vectors (listed in Table 1) with any degree of confidence. Using a uniform distribution, a maximum-entropy Patterson map consistent with the observed data was obtained in 64 iterations. The evolution of the entropy and χ^2 is presented in Fig. 3. The Patterson maps (Fig. 2*d–g*) obtained using the maximum-entropy algorithm contrast markedly with the Patterson maps derived by direct Fourier transformation. Sharp positive and negative features are located close to the correct

TABLE 1 Data from maximum-entropy Patterson map

	x ($\text{\AA}$)	y ($\text{\AA}$)	z ($\text{\AA}$)	Δ ($\text{\AA}$)	ME intensity	True intensity
Peak	0.0000	0.0000	0.0000	0.0000	1.671	1.147
Ti–O	0.3041	0.3041	0.0000	0.0010	–0.387	–0.399
O–O	0.3949	0.3949	0.0000	0.0064	0.325	0.337
Ti–Ti	0.5000	0.5000	0.5000	0.0000	0.166	0.118
Ti–O	0.1942	0.1942	0.5000	0.0014	–0.358	–0.399
O–O	0.5000	0.1110	0.5000	0.0014	0.667	0.673

Comparison of correct and observed (maximum-entropy Patterson map) interatomic vector positions and intensities for TiO_2

Δ , fractional difference between ME and true interatomic vectors

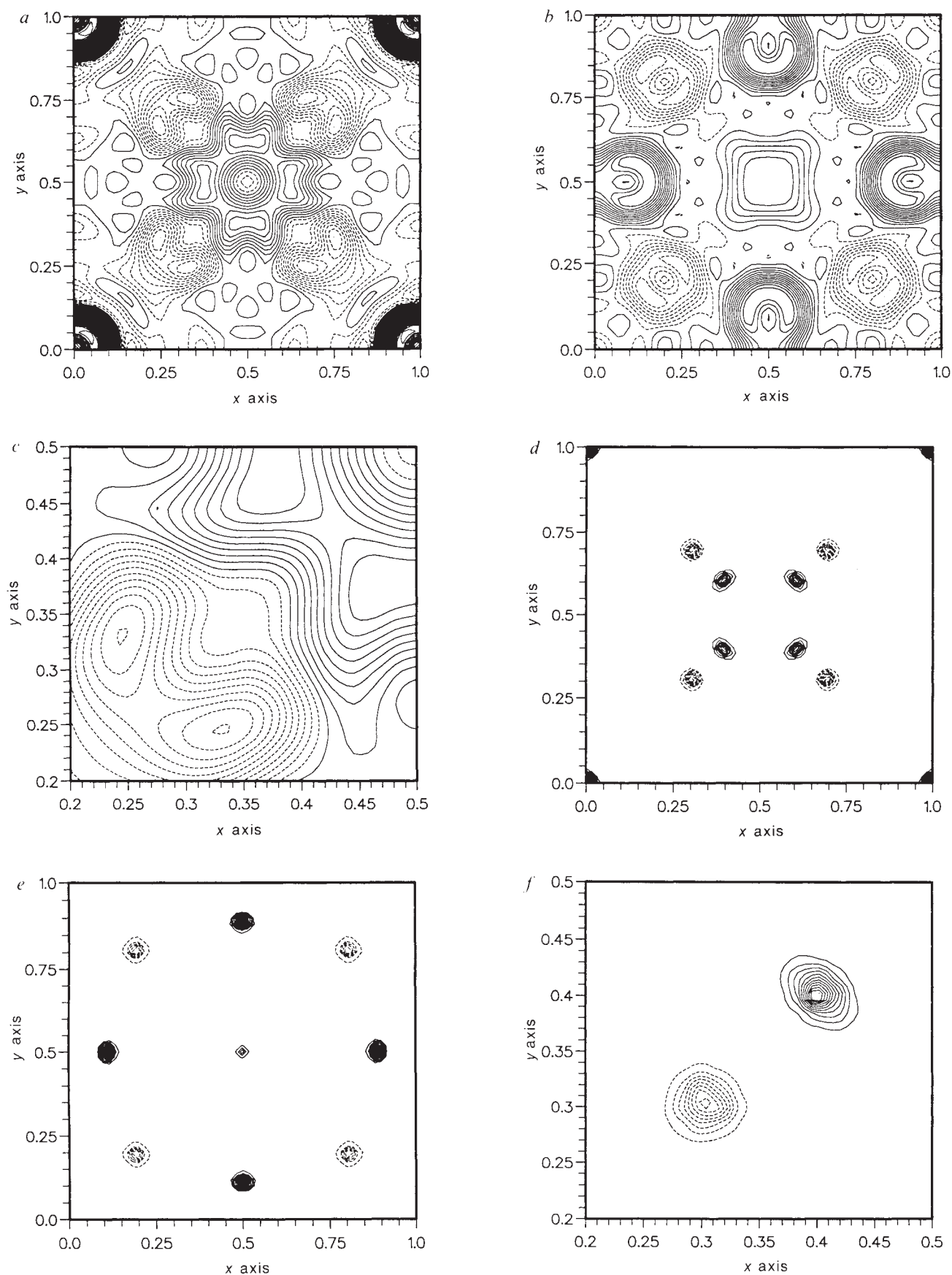


FIG. 2 *a*–*c*, Sections of the Patterson map synthesized by Fourier transformation of the squares of the structure factor magnitudes obtained from the Pawley-type profile refinement of TiO_2 and subsequent averaging over a sphere of diameter 0.5 Å. The first two maps range over the full *xy* plane

at a , $z=0$ and b , $z=\frac{1}{2}$. *c* is a magnification of *a* for $0.2 < x < 0.5$ and $0.2 < y < 0.5$. *d*–*f*, Sections of the Patterson map obtained by maximum entropy reconstruction from the observed clumps of Bragg intensities. The ranges in *d*–*f* correspond to those of *a*–*c*.

interatomic vector positions. Table 1 shows that the typical deviation between observed and correct vector positions is $\sim 0.0015a = 0.007 \text{ \AA}$, a full order of magnitude less than the pixel resolution ($a/60 = 0.08 \text{ \AA}$) and almost two orders of magnitude less than the minimum d spacing measured ($0.487 \text{ \AA}$). The oxygen fractional coordinate calculated from the maximum-entropy Patterson map is 0.3047, essentially identical to the correct value of 0.3048. Moreover, the intensities of these interatomic vectors are in good agreement with expected values (Table 1) and there is even some evidence of the anisotropy of the thermal motion of the oxygen atoms (Fig. 2f). The algorithm has produced an optimally sharpened Patterson map and thus opens up a productive route for the solution of crystal structures.

Evaluation of the individual squared magnitudes of the structure factor was performed by back-transformation of the maximum-entropy Patterson map (for a full list of structure factors see Supplementary Information). A comparison of the following R factors shows that there is a clear superiority in the $|F(\mathbf{h})|^2$ derived from the maximum-entropy algorithm over the *a priori* equipartitioned values.

$$R_{\text{EQ}} = \left[\sum_{\mathbf{h}} w(\mathbf{h}) (F(\mathbf{h})_{\text{EQ}} - F(\mathbf{h}))^2 / \sum_{\mathbf{h}} w(\mathbf{h}) F(\mathbf{h})^2 \right]^{1/2} = 34.0\%$$

$$R_{\text{ME}} = \left[\sum_{\mathbf{h}} w(\mathbf{h}) (F(\mathbf{h})_{\text{ME}} - F(\mathbf{h}))^2 / \sum_{\mathbf{h}} w(\mathbf{h}) F(\mathbf{h})^2 \right]^{1/2} = 14.1\%$$

$$R_{\text{EX}} = \left[N_F / \sum_{\mathbf{h}} w(\mathbf{h}) F(\mathbf{h})^2 \right]^{1/2} = 6.5\%$$

R_{EQ} and R_{ME} are the R factors for the $|F(\mathbf{h})|$ obtained from the refinement of the Pawley profile and from the maximum-entropy algorithm, respectively, and R_{EX} is the expected R factor based on statistical errors. N_F is the number of different $|F(\mathbf{h})|$ and $w(\mathbf{h})$ is the inverse square of the estimated standard deviation of $|F(\mathbf{h})|$. Moreover, for all clumps of reflections the method correctly discriminates between strong and weak intensities, albeit usually with a slightly reduced contrast than for the true values. Indeed, the $|F(\mathbf{h})|^2$ derived from the maximum-entropy algorithm for intensity clumps containing up to five reflections agree well with true values even at short d spacings close to the limits of the diffraction data. A further indication of the success with which the maximum-entropy algorithm has optimally combined the intensity information is provided by consideration of the sign of $|F(\mathbf{h})|^2$. Although the

least-squares profile refinement has yielded negative values for some $|F(\mathbf{h})|^2$, all 177 $|F(\mathbf{h})|^2$ obtained by back-transformation of the maximum-entropy Patterson map are positive quantities. In general, therefore, the $|F(\mathbf{h})|^2$ obtained by the maximum-entropy method can be used to increase the potential for success of a direct-methods approach to structure solution using powder data.

It is worth considering two broader implications of the maximum-entropy formalism that may have important ramifications for *ab initio* structure solution. The Patterson maps produced may be regarded as being optimally sharpened and thus be the closest approach to a true interatomic vector map. Although computationally expensive, this has broad implications for Patterson methods and will assist in the current revival⁸ of the technique as an important alternative to direct methods of structure solution^{9,10}. Furthermore, the formalism uses the information content contained in negative scattering factors. Many aspects of direct methods of structure solution from X-ray data rely implicitly on the inherent positivity of the electron scattering density. This poses concerns for neutron diffraction data where the scattering density may be both positive and negative and, indeed, weakens the power of direct methods in such situations. Here, the negative scattering density is a bonus. Consider, for example, a molecule in an asymmetric unit of a unit cell and containing a single atom with a negative scattering length. The negative peaks in the maximum-entropy Patterson map only correspond to vectors between the negative scatterer and all the other atoms. Such a situation bears a close resemblance to isomorphous replacement^{11,12} and anomalous dispersion¹² methods. Examination of the negative Patterson features uniquely defines the location and orientation of the molecule with respect to the negatively scattering atom. $\square$

Received 19 April; accepted 26 June 1990.

1. Rietveld, H. M. *J. appl. Crystallogr.* **2**, 65–71 (1969).
2. David, W. I. F. *J. appl. Crystallogr.* **20**, 316–319 (1987).
3. Gull, S. F. & Danielli, G. *J. Nature* **272**, 686–690 (1978).
4. Livesey, A. K. & Skilling, J. *Acta Crystallogr.* **A41**, 113–122 (1985).
5. Brice, G. *Acta Crystallogr.* **A40**, 410–445 (1984).
6. Hull, S. & Mayers, J. *User Guide for the Polaris Powder Diffractometer at ISIS, Rep. RAL-89-118* (Rutherford Appleton Laboratory, Chilton, 1989).
7. Pawley, G. S. *J. appl. Crystallogr.* **14**, 347–361 (1981).
8. Egert, E. *Acta Crystallogr.* **A39**, 936–940 (1983).
9. Woolfson, M. M. *Acta Crystallogr.* **A43**, 593–612 (1987).
10. Ladd, M. F. C. & Palmer, R. A. (eds) *Theory and Practice of Direct Methods in Crystallography* (Plenum, London, 1980).
11. Rossmann, M. G., Arnold, E. & Vriend, G. *Acta Crystallogr.* **A42**, 325–334 (1986).
12. Karle, J. *Acta Crystallogr.* **A45**, 765–781 (1989).

SUPPLEMENTARY INFORMATION: Requests should be addressed to the London editorial office of *Nature*.

Sensitivity of regional climates to localized precipitation in global models

A. J. Pitman, A. Henderson-Sellers & Z.-L. Yang

School of Earth Sciences, Macquarie University, North Ryde, NSW 2109, Australia

ATMOSPHERIC general circulation climate models are valuable tools for investigating the effects that large-scale perturbations (for example, increasing greenhouse gases¹, volcanic dust in the atmosphere², deforestation^{3,4} and desertification⁵) might have on climate. Designed for global-scale climate simulations and having coarse spatial resolutions (from $\sim 250 \text{ km}$ (ref. 6) to 800 km (ref. 7)) such models are now being used to evaluate climatological and hydrological quantities at or near the land surface and at sub-continental scales^{8–10}. They are also being used to provide input data for high-resolution simulations, predicting, for example, the

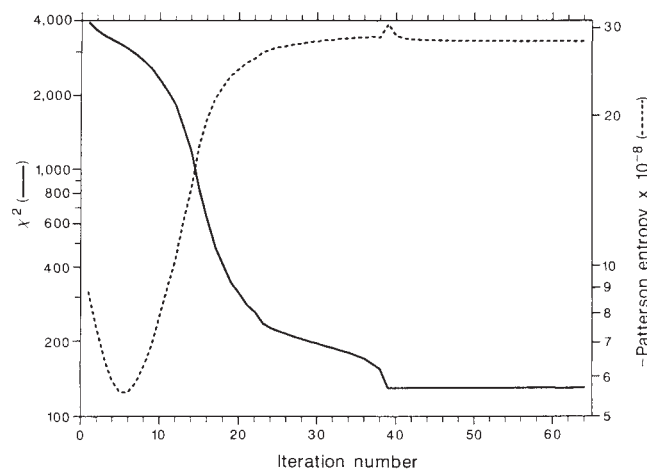


FIG. 3 Evolution of the entropy and χ^2 goodness of fit for the construction of the maximum-entropy Patterson map. (The negative value of the Patterson entropy is plotted as it is an intrinsically negative property.) The entropy decreases initially as the Patterson map changes from the flat uniform distribution to one consistent with the observed data ($\chi^2 = 130$) after 39 iterations. Once agreement with the observed diffraction intensities has been obtained the entropy increases to its maximum value consistent with the χ^2 constraint.

effects of increasing atmospheric CO_2 at regional scales¹¹. Here we show that continental surface climatologies and climate change predictions for tropical forest ecotypes derived from general circulation models may be very sensitive to slight modifications in the land-surface/atmosphere coupling. Specifically our results indicate that improving the realism of the areal distribution of precipitation alters the balance between runoff and evaporation. In the case of a tropical forest, we find that by modifying the area over which rainfall is distributed, the surface climatology is changed from an evaporation-dominated regime to one dominated by runoff. Climate studies using near-surface output from such simulations may therefore be misleading.

The land surface is an essential component in atmospheric general circulation models (AGCMs). A series of sensitivity experiments¹² has shown that the simulation of the Earth's climate is affected by the land surface^{3,8} and its representation^{9,13}. In the past decade several models for parameterizing the land surface in AGCMs have been designed^{4,14,15}. The latest include detailed soil and canopy models which, under the observed climatological forcing, realistically simulate surface-atmosphere energy fluxes, temperatures and the water balance for specific locations^{3,9}. Although these land surface models are realistic, the atmospheric quantities output by the AGCM are not always compatible with those required by the land surface models. For instance, virtually all current AGCMs predict the precipitation rate as a grid-element average, that is, a single value for precipitation is calculated for each grid element at each timestep. Thus both large-scale (such as frontal) and small-scale (such as convective) precipitation is assumed to fall uniformly over the entire grid element. In the case of large-scale precipitation this may be satisfactory (particularly for models with spatial resolutions of $\sim 300 \times 300$ km). In the case of small-scale precipitation, however, this grid-average assumption leads to a loss of physical realism. Specifically, current AGCMs predict drizzle across entire grid areas rather than intense convective precipitation concentrated in much smaller areas within grids. In the former situation a much greater proportion of the rainfall will be intercepted by the vegetation canopy and subsequently re-evaporated, whereas in the latter, more realistic situation, throughfall and hence runoff will be much greater. This modification in the partitioning of precipitation between runoff and evaporation has the potential to affect the energy and hydrological balance of the land surface. It may also cause systematic changes in the predictions by AGCMs once feedbacks are incorporated. These effects will be most pronounced in regions where dense vegetation exists and where convective precipitation is common, for example, in tropical forest areas.

Tropical convective precipitation can produce point-measured precipitation intensities of 10 mm h^{-1} but the spatial extent of such storms is only $\sim 100 \text{ km}^2$. At this intensity, a canopy would saturate rapidly leading to leaf drip and therefore a large proportion of the precipitation would reach the soil surface where it would either infiltrate or run off. In an AGCM, convective precipitation is spread over the entire grid element. Assuming grid areas of $100,000 \text{ km}^2$ ($\sim 3^\circ \times 3^\circ$) and a precipitation intensity of 10 mm h^{-1} leads to a precipitation intensity over the entire grid element of $\sim 10^{-2} \text{ mm h}^{-1}$. In this case the canopy would intercept and retain virtually all the precipitation, contrary to observation.

The evaporation of precipitation intercepted by a canopy occurs very rapidly in comparison to the evaporation of water that infiltrates the soil surface. The low precipitation intensities 'predicted' by AGCMs could therefore lead to fundamentally misleading hydrological simulations because precipitation recycling from the surface to the atmosphere, in any AGCM incorporating a vegetation canopy, will be overestimated at the expense of increasing soil moisture or runoff. To overcome this problem it is important to account for both the spatial extent and intensity of precipitation in AGCMs that incorporate the new generation of land surface schemes.

Here we investigate, using the Biosphere-Atmosphere Transfer Scheme (BATS)¹⁵, a method of retaining the spatial extent and intensity of precipitation for AGCMs suggested elsewhere^{16,17}. Throughout this discussion water fluxes are in units of $\text{kg m}^{-2} \text{ s}^{-1}$ ($1 \text{ kg m}^{-2} \text{ s}^{-1} = 1 \text{ mm s}^{-1}$). It is possible to derive an expression for surface runoff¹⁶ R_{surf} and canopy drip¹⁷ R_{drip} by assuming that the local precipitation rate (over a fraction μ of the grid element) is described by a decaying exponential probability distribution. We note that although subgridscale variations exist in R_{surf} and R_{drip} within a timestep, knowledge of the distribution of local precipitation is not propagated—at the end of a timestep the soil moisture and intercepted water are assumed to be distributed uniformly within the grid element. This is clearly inconsistent with the definition of the net flux of water at the surface, P_s , but no method yet exists to account for subgridscale variations in intercepted water and soil moisture in AGCMs. Expressions for R_{surf} and R_{drip} are given by

$$R_{\text{surf}} = P_s \exp\left(\frac{-\mu F_s}{P_s}\right) \quad (1)$$

where F_s is the maximum surface infiltration rate, assumed constant over the grid element. Similarly, the canopy drip rate is

$$R_{\text{drip}} = P_c \exp\left(\frac{-\mu F_c}{P_c}\right) \quad (2)$$

where P_c is the precipitation intercepted by the canopy and F_c is the maximum canopy infiltration rate, assumed constant over the grid element. F_c is defined as the difference between the canopy storage capacity and the water stored on the canopy, divided by the model timestep.

In the standard version of BATS, R_{surf} is given by

$$R_{\text{surf}} = \beta^4 P_s \quad (3)$$

where β is a depth-weighted ratio of the soil wetness. Equation (3) predicts $R_{\text{surf}} > 0$ when any water falls to the surface. The canopy drip is usually determined by the storage capacity S (kg m^{-2}). In a timestep Δt when the mass of intercepted water per unit leaf area C (kg m^{-2}) on the canopy exceeds S , all water in excess of S falls to the surface

$$R_{\text{drip}} = (C - S)/\Delta t \quad (4)$$

Here we replace these standard formulae by equations (1) and (2).

Using the modified version of BATS in a stand-alone mode¹⁰ (that is, uncoupled from the GCM), and prescribing atmospheric forcing (including precipitation, air temperature, solar radiation, downward near-infrared radiation and wind speed) with seasonal and diurnal cycles from climatological estimates¹⁸, the model was integrated for a two-year simulation to investigate the sensitivity of the land surface to changes in the definition of μ . Here we report results of the second year of integration (these results are independent of initial conditions) for a tropical forest, with μ set to 1.0 (rainfall over the whole grid element), 0.5 and 0.1. The results from the control simulation using the standard version of BATS are also included. We consider tropical forest and tropical rainfall because the dense canopy and the high leaf-area index typical of this ecotype leads to a high sensitivity to the type (and here the spatial extent) of precipitation simulated by AGCMs and because of the number of Amazonian deforestation experiments conducted using AGCMs^{3,4,19}. These results are, however, likely to be appropriate to all regions where precipitation falls at spatial scales less than the resolution of the AGCM.

Figure 1 shows four sets of histograms of precipitation, evaporation and runoff for the four simulations. In each case the monthly total precipitation is identical, as is the atmospheric forcing but the fraction of the grid square over which the precipitation is prescribed to fall is variable.

The control simulation using the standard BATS model (Fig. 1a) shows that the seasonal variation in evaporation is limited,

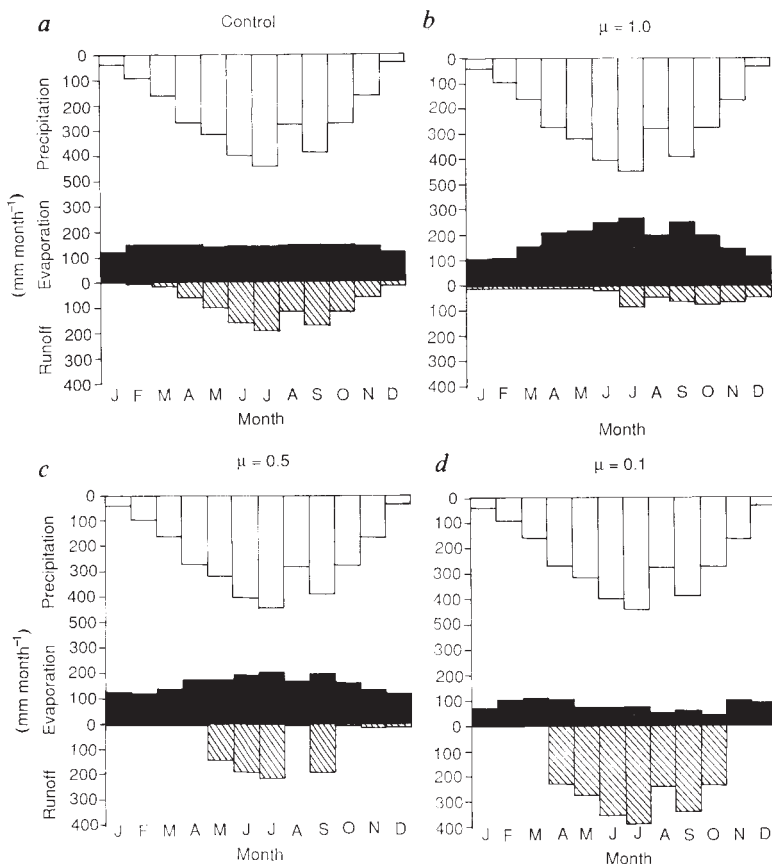


FIG. 1 Results from the second year of four 2-yr stand-alone simulations showing the seasonal variation in the precipitation (open bars), evaporation rates (solid bars) and runoff (hatched bars) for *a*, the control simulation (standard BATS without the μ -parameterization), *b*, BATS plus the μ -parameterization with $\mu = 1.0$; *c*, $\mu = 0.5$; *d*, $\mu = 0.1$. The precipitation forcing varies from 34 mm month⁻¹ in December to 440 mm month⁻¹ in July. The air temperature varies seasonally and diurnally, with the annual maximum in March–April (301 K) and the minimum in July–August (298 K) and a diurnal temperature range of 3 K. The wind velocity (at the height of the canopy) is constant at 3 m s⁻¹ and the atmospheric pressure is constant at 1,000 hPa. Solar radiation varies seasonally and diurnally and is reduced after rainfall in an attempt to simulate the gross effects of cloud cover.

with the precipitation in excess of evaporation during the summer months (JJA) forming runoff. BATS incorporating the μ -parameterization (replacing equations (3) and (4) by equations (1) and (2)) produces very different simulations. Figure 1*b* ($\mu = 1.0$) shows much higher evaporation fluxes and very little runoff. For $\mu = 1.0$, the precipitation is distributed over the entire grid square leading to relatively low intensities. The dense canopy of the tropical forest intercepts much of the rainfall leading to higher monthly evaporation, but much lower runoff rates.

Although the spatial extent of precipitation in Fig. 1*a, b* is the same (distributed over the entire grid element, although non-uniform in Fig. 1*b*) it is clear that the resulting simulation of evaporation and runoff is quite different because the actual formulations of R_{surf} and R_{drip} are different. This leads to a change in the amount of precipitation that is intercepted and evaporated or that reaches the soil surface. The differences between Fig. 1*a, b* show how sensitive the simulation of the partitioning of precipitation between runoff and evaporation is to minor changes in the formulation of the land surface in AGCMs.

As μ is reduced from 1.0, the effective precipitation intensity is increased and interception rates are reduced, leading to lower evaporation and higher runoff. Figure 1*c* shows the case for $\mu = 0.5$. Evaporation shows some dependence on season (although less than in Fig. 1*b*) and high runoff during the summer months. Figure 1*c* shows negligible runoff in April, October and November, however, in contrast to Fig. 1*a*. Finally, Fig. 1*d* shows the simulation for $\mu = 0.1$. Here, precipitation is concentrated over 10% of the grid element, hence its effective intensity is relatively high. Figure 1*d* shows that there is relatively little evaporation except in February, March, April, November and December when evaporation rates are comparable with the minimum rates in Fig. 1*a–c*. The runoff simulated with $\mu = 0.1$ is dramatically higher than in the previous simulations. The high precipitation intensities lead to very low interception and high runoff rates for seven months of the year. In this case the canopy

becomes saturated rapidly in a precipitation event, leading to high throughfall and subsequently large runoff rates.

The change from Fig. 1*a* or Fig. 1*b* to Fig. 1*d* is somewhat similar (decreased evaporation and increased runoff) to modelled results of tropical deforestation³ although the only disturbance is in the spatial distribution of a constant precipitation amount falling on unchanged tropical forest. This implies that recent estimates of the impact on climate of tropical deforestation^{3,4,19} are inconsistent because two^{4,19} include some description of subgridscale precipitation and two^{3,19} include a canopy parameterization. It is probably unrealistic simply to incorporate a μ -parameterization without considering very carefully the likely impact on the hydrology, which is a function of the canopy scheme. The relative and combined importance of subgridscale precipitation and vegetation must be more thoroughly investigated.

The quantitative values in the histograms of runoff and evaporation in Fig. 1 are of less importance (because minor changes in the land surface parameterization or in the prescribed atmospheric forcing would alter magnitudes) than the relative shapes of the seasonal distributions. The differences in the seasonal distributions shown in Fig. 1*a–d* are of considerable importance to potential users of AGCM simulations for quantities at or near the land surface. Figure 1 shows clearly that the surface hydrological climatology is highly sensitive to atmospheric 'forcing', in this case precipitation. Simply by changing the area over which precipitation is distributed, the surface climatology can be changed from an evaporation-dominated regime (Fig. 1*b*) to one dominated by runoff (Fig. 1*d*).

By prescribing atmospheric forcing, feedback effects between the surface and the atmosphere are prevented. As it is not obvious what direction feedbacks would take, it is unclear whether the simulations presented here are amplifications of the fully interactive result, or whether the results are damped. It is therefore necessary to analyse the μ -type parameterization in an AGCM before developing the methodology further.

These simulations suggest that it may not be possible to

improve the simulation of climate at the land surface by improving only the land surface models. Advances in land surface modelling must be made in tandem with improvements in the coupling between the land surface and atmosphere. The addition of vegetation and soil processes to AGCMs improves physical realism, but it also has the potential to increase the sensitivity of the surface to the atmosphere. If incorporation of these more complete land surface submodels is considered desirable then the simulation of precipitation, including its subgridscale variability, must be improved. We do not necessarily claim improved accuracy for any of the regimes illustrated in Fig. 1, because observational evidence is not yet sufficiently accurate to validate any particular simulation. Rather we point out that including the effects of spatial heterogeneity in precipitation increases the physical realism of the model simulations and therefore its consequences must be fully investigated.

The distribution of precipitation within grid elements needs to be incorporated into the AGCMs that include vegetation. Its calculation is fairly straightforward²⁰ but even specifying precipitation intensities of large-scale and small-scale precipitation events as different but constant values^{16,17} would improve the simulation of near-surface quantities. This approach is dangerous, however, because a constant value for μ could lead to a climatology dependent on the value chosen for μ . Figure 1a-d implies that using AGCM simulations of near-surface variables from models that incorporate a parameterization of vegetation, but do not consider subgridscale variability of

precipitation, are misleading. Our results indicate that using results of AGCMs for regional-scale impact studies of, for example, greenhouse warming or deforestation might be very misleading unless the limitations of the model are clearly understood. □

Received 16 March; accepted 9 July 1990.

1. Mitchell, J. F. B., Wilson, C. A. & Cunningham, W. M. *Q. J. R. met. Soc.* **113**, 293–322 (1987).
2. Hansen, J. *et al. Science* **213**, 957–966 (1981).
3. Dickinson, R. E. & Henderson-Sellers, A. *Q. J. R. met. Soc.* **114**, 439–462 (1988).
4. Lean, J. & Warrilow, D. A. *Nature* **342**, 411–413 (1989).
5. Charney, J., Quirk, W. J., Chow, S.-H. & Kornfeld, J. *J. Atmos. Sci.* **34**, 1366–1385 (1977).
6. Rowntree, P. R. & Bolton, J. A. *Q. J. R. met. Soc.* **109**, 501–526 (1983).
7. Hansen, J. E. *et al. Mon. Wea. Rev.* **111**, 609–662 (1983).
8. Rind, D. *J. geophys. Res.* **93**, 5385–5412 (1988).
9. Sato, N. *et al. NASA Contractor, Rep. 185509* (NASA, Washington, DC, 1989).
10. Wilson, M. F., Henderson-Sellers, A., Dickinson, R. E. & Kennedy, P. J. *J. Clim. appl. Met.* **26**, 341–362 (1987).
11. Cohen, S. J. & Allsopp, T. R. *J. Clim.* **1**, 669–681 (1988).
12. Mintz, Y. in *The Global Climate* (ed. Houghton, J. T.) 79–105 (Cambridge University Press, 1984).
13. Mitchell, J. F. B. & Warrilow, D. A. *Nature* **330**, 238–240 (1987).
14. Sellers, P. J., Mintz, Y., Sud, Y. C. & Dalcher, A. *J. Atmos. Sci.* **43**, 505–531 (1986).
15. Dickinson, R. E., Henderson-Sellers, A., Kennedy, P. J. & Wilson, M. F. *NCAR Tech. Note, TN275 + STR* (Natl. Center for Atmospheric Research, Boulder, 1986).
16. Warrilow, D. A., Sangster, A. B. & Slingo, A. *Dynamic Climatology Tech. Note 38* (Meteorological Office, Bracknell, 1986).
17. Shuttleworth, W. J. *J. Hydrol.* **100**, 31–56 (1988).
18. Dickinson, R. E. (ed.) *The Geophysics of Amazonia* (Wiley, New York, 1987).
19. Shukla, J., Nobre, C. & Sellers, P. J. *Science* **247**, 1322–1325 (1990).
20. Entekhabi, D. & Eagleson, P. S. *J. Clim.* **2**, 816–831 (1989).

ACKNOWLEDGEMENTS. We thank R. E. Dickinson and W. J. Shuttleworth for their comments on the manuscript. A.J.P. acknowledges a Macquarie University Research Fellowship. This work was funded in part by the Australian Research Council.

Partly pedogenic origin of magnetic variations in Chinese loess

L. P. Zhou*, F. Oldfield†, A. G. Wintle‡, S. G. Robinson§ & J. T. Wang||

* Godwin Laboratory, Subdepartment of Quaternary Research, University of Cambridge, Free School Lane, Cambridge CB2 3RS, UK

† Department of Geography, University of Liverpool, PO Box 147, Liverpool L69 3BX, UK

‡ Institute of Earth Studies, University College of Wales, Aberystwyth SY23 3DB, UK

§ Department of Earth Sciences, University of Cambridge, Downing Street, Cambridge CB2 3EQ, UK

|| Geological Hazards Research Institute, Lanzhou, Gansu, China

QUATERNARY sequences of interbedded aeolian silts (loess) and buried soils (palaeosols) potentially provide one of the best terrestrial records of past climates^{1–5}. Magnetic susceptibility variations in loess and palaeosol sequences from China are strongly correlated with climate-induced fluctuations of oxygen isotope ratios in deep-sea sediments^{2–4,6–8}. As an explanation of this correlation, it has been suggested that the recorded variations in magnetic susceptibility depend primarily on the degree of dilution by non-magnetic bulk loess constituents of a uniform magnetic assemblage derived from remote but unknown sources^{6–8}. It has also been suggested that such a model, involving changes in the accumulation rate of loess superimposed on a constant flux of magnetic particles, provides a relative geological timescale through the control of loess deposition by astronomically modulated climate fluctuations⁷. Here we report a preliminary test of this model using rock magnetic properties especially sensitive to variations in magnetic grain size. Our results show that palaeosols are characterized by much finer magnetic grain size assemblages than are the intervening loess units. This suggests that a simple model based on constant magnetic influx and dilution by variable amounts of non-magnetic loess is inadequate. Our magnetic measurements establish the close comparability of the fine grained magnetic minerals in the

palaeosol samples to those in contemporary soils and thus point to a partially pedogenic origin for the magnetic mineral assemblages in the palaeosols.

Here we present results obtained from two groups of samples. One (Fig. 1) consists of 17 samples of late Pleistocene loess and a palaeosol from Xifeng, Gansu (35.7° N, 107.6° E). The other (Fig. 2a, b) consists of samples taken from eight successive palaeosols and the loess units above and below them in the upper part of a Quaternary loess section near Lanzhou, Gansu (36.1° N, 103.8° E). Magnetic measurements were made on air-dried, disaggregated and weighed subsamples.

Several recent studies describe the instrumentation used for magnetic measurements and outline the way in which they may be interpreted^{9–11}. These studies show how rock magnetic measurements, which reflect variations in magnetic mineralogy and grain size, can be used to characterize soils, dusts and sediments, to differentiate them and to establish their origins. Here the results from Xifeng (Fig. 1) are plotted in stratigraphic sequence, whereas those from the Lanzhou section (Fig. 2a, b) are presented as bivariate plots.

Figure 1 shows that field and laboratory measurements of the profile of the magnetic susceptibility are in excellent agreement, and that the latter are only slightly affected by expressing the results on a carbonate-free basis. The magnetic susceptibilities in palaeosol horizons are over three times higher and the χ_{fd} and χ_{ARM} values, which are strongly correlated, generally five to seven times higher than those in the loess. Moreover, variations in the normalized quotients $\chi_{fd}(\%)$ and $\chi_{ARM}/SIRM$ are positively correlated and in $SIRM/\chi$ negatively correlated with those in bulk magnetic susceptibility. These variations must therefore reflect differences between the samples in the relative proportions of different magnetic components, and cannot be caused simply by changes in the degree of dilution by non-magnetic constituents in the samples. More specifically, variations in $\chi_{ARM}/SIRM$ can be explained only in terms of the behaviour of the remanence-carrying minerals and cannot be affected by paramagnetic or diamagnetic components.

Measurements of IRM acquisition (not shown here) confirm that all samples from both sections attain 92–97% of the SIRM in a forward field of 300 mT. Therefore we conclude that the

magnetic properties of these samples are dominated by ferri-magnetic (for example, magnetite-type) behaviour and that variations between the samples in the relative importance of imperfect anti-ferromagnets (such as haematite) will not alone account for the large differences in rock magnetic properties encountered.

Many studies have shown that variations in ferrimagnetic grain size and thus domain structure have an important influence on the range of the magnetic parameters we have measured¹²⁻¹⁶. Maher¹⁷ has demonstrated that in dispersed very-fine-grained synthetic magnetite samples with grain-interaction characteristics comparable to those in natural samples, two normalized properties, $\chi_{fd}(\%)$ and $\chi_{ARM}/SIRM$, are especially responsive to variations in the magnetic grain size. Although based on simple assemblages of synthetic magnetite, her observations are consistent with a wide range of data from natural samples which include more complex mineral assemblages^{9-11,18-21}.

In our data, these two properties show a high degree of correlation. Palaeosol samples are characterized by $\chi_{fd}(\%)$ values up to 8-10%, and $\chi_{ARM}/SIRM$ values up to $60-80 \times 10^{-5} A^{-1} m$. The intervening loess samples have lower values; $\chi_{fd}(\%)$ is rarely over 5% and $\chi_{ARM}/SIRM$ is rarely over $35 \times 10^{-5} A^{-1} m$.

In Maher's New MT series of synthetic samples¹⁷, maximum $\chi_{fd}(\%)$ values of 7-11% occur in samples with grain sizes around the transition from stable single domain (SSD) to superpara-magnetic (SP) behaviour, and only samples with mean grain diameters of 0.016-0.036 μm yield values over 2.5%. Peak values of $\chi_{ARM}/SIRM$ occur in the true SSD range, with values over $60 \times 10^{-5} A^{-1} m$ only in samples with mean grain diameters

$<0.07 \mu m$. Magnetic grain assemblages that span the SP/SSD transition are the only ones with high values for both parameters.

From these observations we infer that the palaeosol horizons in both sections are relatively enriched in ferrimagnetic grains around the SP/SSD transition. These grains are finer than those responsible for the magnetic properties of the intervening loess. The recorded variations in $SIRM/\chi$ reflect the same shifts in magnetic grain size. The reciprocal relationship between $\chi_{ARM}/SIRM$ and $\chi_{fd}(\%)$ on the one hand (Fig. 2a), and $SIRM/\chi$ on the other (Fig. 2b), arises because grains with sizes above the SP/SSD boundary have relatively low χ and high SIRM, whereas true SP grains exhibit high χ and zero SIRM. Lower $SIRM/\chi$ quotients are thus indicative of the finer grains, and vice versa. The magnetic properties of the palaeosols compare well with those of many contemporary soils. For example, the majority of $\chi_{fd}(\%)$ and $\chi_{ARM}/SIRM$ values for soils developed on contrasted lithologies in the Chesapeake Bay area¹¹ range from 7% to 13% and from 40 to $120 \times 10^{-5} A^{-1} m$ respectively. We conclude that significant proportions of the fine grained ferrimagnets characteristic of these palaeosols are pedogenic in origin and have been formed *in situ*.

Magnetic enhancement in topsoil has long been recognized²²⁻²⁵. The effects of fire and the *in situ* formation of maghemite and/or magnetite under soil-forming conditions are thought to be the dominant mechanisms involved^{9,22-25}. Recent studies have demonstrated that in both natural and simulated soil environments magnetite can be formed through inorganic processes^{26,27}. Meanwhile magnetotactic bacteria²⁸ have been found in modern soils²⁹. The secondary magnetic phases, however formed in soils, have been shown to survive for long

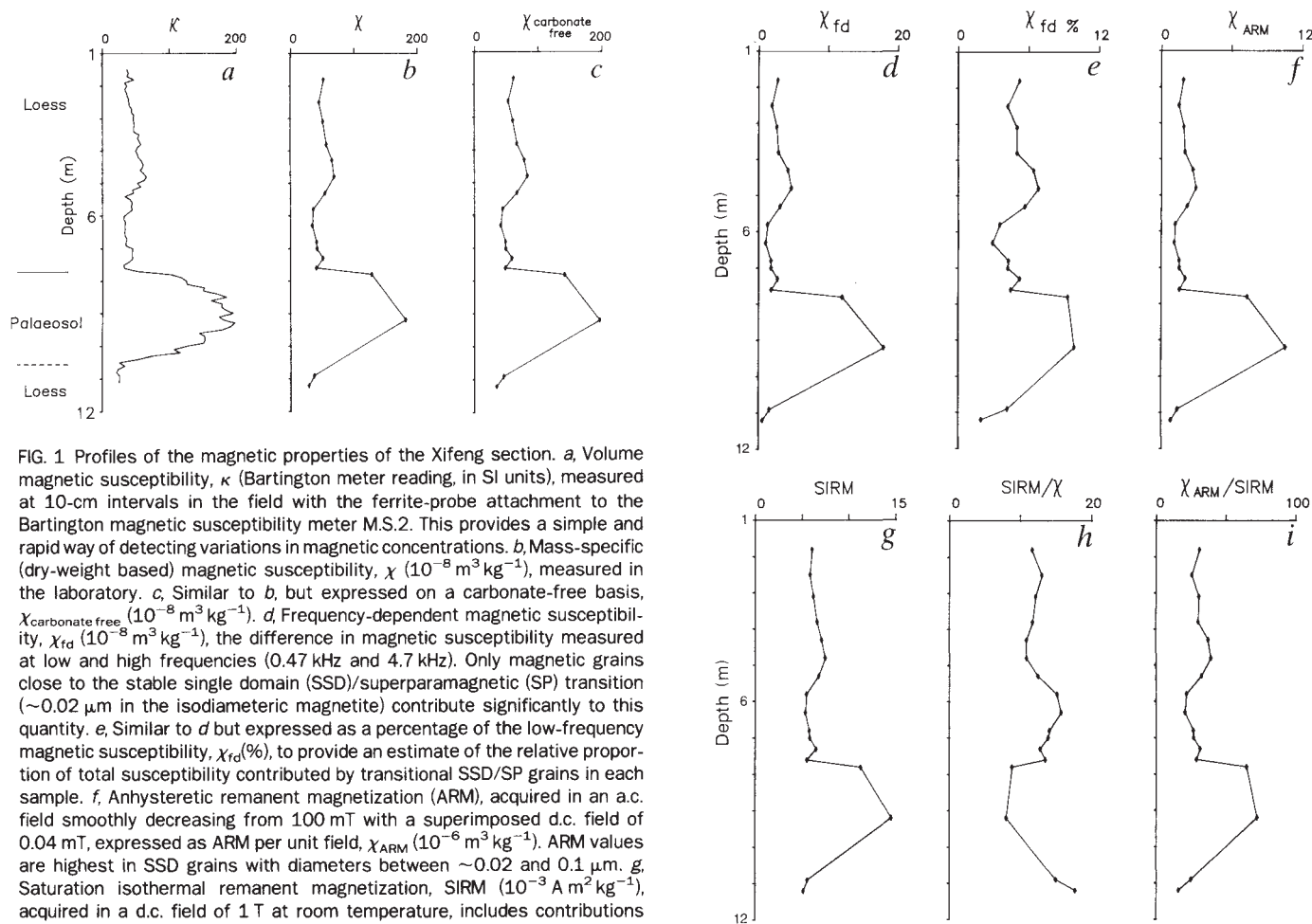


FIG. 1 Profiles of the magnetic properties of the Xifeng section. *a*, Volume magnetic susceptibility, κ (Bartington meter reading, in SI units), measured at 10-cm intervals in the field with the ferrite-probe attachment to the Bartington magnetic susceptibility meter M.S.2. This provides a simple and rapid way of detecting variations in magnetic concentrations. *b*, Mass-specific (dry-weight based) magnetic susceptibility, χ ($10^{-8} m^3 kg^{-1}$), measured in the laboratory. *c*, Similar to *b*, but expressed on a carbonate-free basis, $\chi_{carbonate free}$ ($10^{-8} m^3 kg^{-1}$). *d*, Frequency-dependent magnetic susceptibility, χ_{fd} ($10^{-8} m^3 kg^{-1}$), the difference in magnetic susceptibility measured at low and high frequencies (0.47 kHz and 4.7 kHz). Only magnetic grains close to the stable single domain (SSD)/superparamagnetic (SP) transition ($\sim 0.02 \mu m$ in the isodiametric magnetite) contribute significantly to this quantity. *e*, Similar to *d* but expressed as a percentage of the low-frequency magnetic susceptibility, $\chi_{fd}(\%)$, to provide an estimate of the relative proportion of total susceptibility contributed by transitional SSD/SP grains in each sample. *f*, Anhyseretic remanent magnetization (ARM), acquired in an a.c. field smoothly decreasing from 100 mT with a superimposed d.c. field of 0.04 mT, expressed as ARM per unit field, χ_{ARM} ($10^{-6} m^3 kg^{-1}$). ARM values are highest in SSD grains with diameters between ~ 0.02 and $0.1 \mu m$. *g*, Saturation isothermal remanent magnetization, SIRM ($10^{-3} A m^2 kg^{-1}$), acquired in a d.c. field of 1 T at room temperature, includes contributions from all the remanence-carrying magnetic minerals in the sample. *h*, $SIRM/\chi$ ($kA m^{-1}$). *i*, $\chi_{ARM}/SIRM$ ($10^{-5} A^{-1} m$). The χ_{ARM} and χ_{fd} variations confirm the relatively high concentrations of fine SSD and SP grains in the palaeosol,

the $\chi_{fd}(\%)$ and $\chi_{ARM}/SIRM$ values establish their greater importance in the palaeosol as a proportion of the total magnetic mineral assemblage.

periods under favourable conditions^{9,25,30,31}. The magnetic susceptibility of loess seems to be mainly, but probably not completely (see below), due to aeolian detrital magnetic minerals of larger mean grain size.

In addition to the contrast in magnetic properties between the palaeosol and loess, there are variations in both concentration-dependent and concentration-normalized properties within each. For example, the $\chi_{fd}(\%)$ for loess is ~2–5% (Fig. 1e) at Xifeng and 0.6–3% (Fig. 2a) at Lanzhou, whereas preliminary studies on samples from a site closer to desert areas to the north yield $\chi_{fd}(\%)$ values rarely above 1% (L.P.Z. unpublished data). This may reflect differences in provenance^{5,19,32}, varying distance from source areas^{5,32} or some *in situ* pedogenic transformation during loess accumulation. Several studies^{9,19} of contemporary dusts and aerosols suggest that variations in the magnetic properties of their primary detrital components are likely to reflect provenance-related differences rather than changes in volcanic or cosmic fluxes. Furthermore, in the Lanzhou samples, there is a wide range of variation in magnetic properties among the palaeosol samples although we observe no overall temporal trend (Fig. 2a, b). The variations may reflect differential magnetic enhancement resulting from differences in past pedo-environments. At sites where parent materials, topography and climate were comparable³³, such a mechanism could give rise to the parallel variations of magnetic susceptibility as previously observed^{6–8}.

In so far as these and more detailed studies in China (B. A. Maher and R. Thompson, manuscript in preparation) and else-

where^{34,35} point to the role of pedogenesis and of possible source-related variations in controlling the magnetic record in loess sections, they replace the idealized constant magnetic influx-dilution model with a more complex one which could give more information about palaeoenvironment once fundamental problems of loess chronology are satisfactorily resolved. □

Received 27 March; accepted 9 July 1990.

1. Fink, J. & Kukla, G. J. *Quat. Res.* **7**, 363–371 (1977).
2. Liu, T. S. et al. *Loess and the Environment* (China Ocean Press, Beijing, 1985).
3. Sasajima, S. & Wang, Y. Y. (eds) *The Recent Research of Loess in China* (Kyoto University & Northwest University, Kyoto, 1984).
4. Heller, F. & Liu, T. S. *Geophys. Res. Lett.* **13**, 1169–1172 (1986).
5. Pye, K. *Aeolian Dust and Dust Deposits* (Academic, London, 1987).
6. Kukla, G. J. *Quat. Sci. Rev.* **6**, 191–219 (1987).
7. Kukla, G. J. et al. *Geology* **16**, 811–814 (1988).
8. Kukla, G. J. & An, Z. *Palaeogeogr. Palaeoclimatol. Palaeoecol.* **72**, 203–225 (1989).
9. Thompson, R. & Oldfield, F. *Environmental Magnetism* (Allen & Unwin, London, 1986).
10. Yu, L. & Oldfield, F. *Quat. Res.* **32**, 168–181 (1989).
11. Oldfield, F., Maher, D. A. & Appleby, P. G. *J. Quat. Sci.* **4**, 189–200 (1989).
12. Parry, L. G. *Phil. Mag.* **11**, 303–312 (1965).
13. Mullins, C. E. & Tite, M. S. *J. geophys. Res.* **78**, 804–809 (1973).
14. Banerjee, S. K. & Moskowitz, B. M. in *Magnetite Biomineralization and Magnetoreception in Organisms* (eds Kirschvink, J. L., Jones, D. S. & MacFadden, B. J.) **17**–41 (Plenum, New York, 1985).
15. Dunlop, D. J. *Phys. Earth planet. Inter.* **26**, 1–26 (1981).
16. Thompson, R. et al. *Science* **207**, 481–486 (1980).
17. Maher, B. A. *Geophys. J. R. astr. Soc.* **94**, 83–96 (1988).
18. Oldfield, F., Maher, B. A., Donoghue, J. & Pierce, J. *J. geol. Soc. Lond.* **142**, 1035–1046 (1985).
19. Oldfield, F. et al. *Nature* **317**, 516–518 (1985).
20. Ozdemir, O. & Banerjee, S. K. *Earth planet. Sci. Lett.* **59**, 393–403 (1982).
21. King, J., Banerjee, S., Marvin, J. & Ozdemir, O. *Earth planet. Sci. Lett.* **59**, 404–419 (1982).
22. Le Borgne, E. *Ann. Geophys.* **11**, 399–419 (1955).
23. Tite, M. S. & Mullins, C. E. *Archaeometry* **13**, 209–219 (1971).
24. Schwertmann, U. *Adv. Soil Sci.* **1**, 171–200 (1985).
25. Maher, B. A. *Phys. Earth planet. Inter.* **42**, 76–92 (1986).
26. Taylor, R. M., Maher, B. A. & Seif, P. G. *Clay Miner.* **22**, 411–422 (1987).
27. Maher, B. A. & Taylor, R. M. *Nature* **336**, 368–370 (1988).
28. Blakemore, R. P. A. *Rev. Microbiol.* **36**, 217–238 (1982).
29. Fassbinder, J. W. E., Stanjek, H. & Vali, H. *Nature* **343**, 161–163 (1990).
30. Fine, P., Singer, M. J., La Ven, R., Verosub, K. & Southard, R. J. *Geoderma* **44**, 287–306 (1989).
31. Heller, F. & Liu, T. S. *Geophys. J. R. astr. Soc.* **77**, 125–144 (1984).
32. Pye, K. & Zhou, L.-P. *Palaeogeogr., Palaeoclimatol., Palaeoecol.* **73**, 11–23 (1989).
33. Zhu, X., Li, Y., Peng, X. & Zhang, S. *Geoderma* **29**, 237–255 (1983).
34. Beget, J. E. & Hawkins, D. *Nature* **337**, 151–153 (1989).
35. Beget, J. E., Stone, D. B. & Hawkins, D. B. *Geology* **18**, 40–43 (1990).

ACKNOWLEDGEMENTS. We thank N. J. Shackleton, B. A. Maher and K. Pye for suggestions, L. Yu, P. L. Gibbard, E. Derbyshire, R. Grün, D. J. Robertson and S. Crowhurst for encouragement, and R. Jude for laboratory assistance. L.P.Z. thanks St John's College, Cambridge, UK Committee of Vice-chancellors and Principals, Lundgren Fund and the Quaternary Research Association for financial support, Bartington Instruments Ltd for promptly supplying equipment for fieldwork and Z. An and D. Sun for assistance in the field. This is publication 120 of the Institute of Earth Studies, University College of Wales (Aberystwyth).

Differential denudation and flexural isostasy in formation of rifted-margin upwarps

A. R. Gilchrist*† & M. A. Summerfield†‡

* Department of Earth Sciences, University of Liverpool, PO Box 147, Liverpool L69 3BX, UK

† Macrogeomorphology Research Group, Department of Geography, University of Edinburgh, Edinburgh EH8 9XP, UK

MARGINAL upwarps are common features of rifted continental margins^{1,2}, but tectonic models of the evolution of rifted margins have not adequately explained their form, or their persistence along some margins more than 100 Myr after continental rapture. Marginal upwarps not only significantly influence the geomorphological evolution of rifted margins and their adjacent continental interiors, but are also important in determining patterns of offshore sedimentation. Here we show that the contrast in denudation rates between the evolving coastal flanks of rifted margins and their interior hinterlands can promote significant marginal upwarps if the lithosphere responds flexurally to the resulting differential

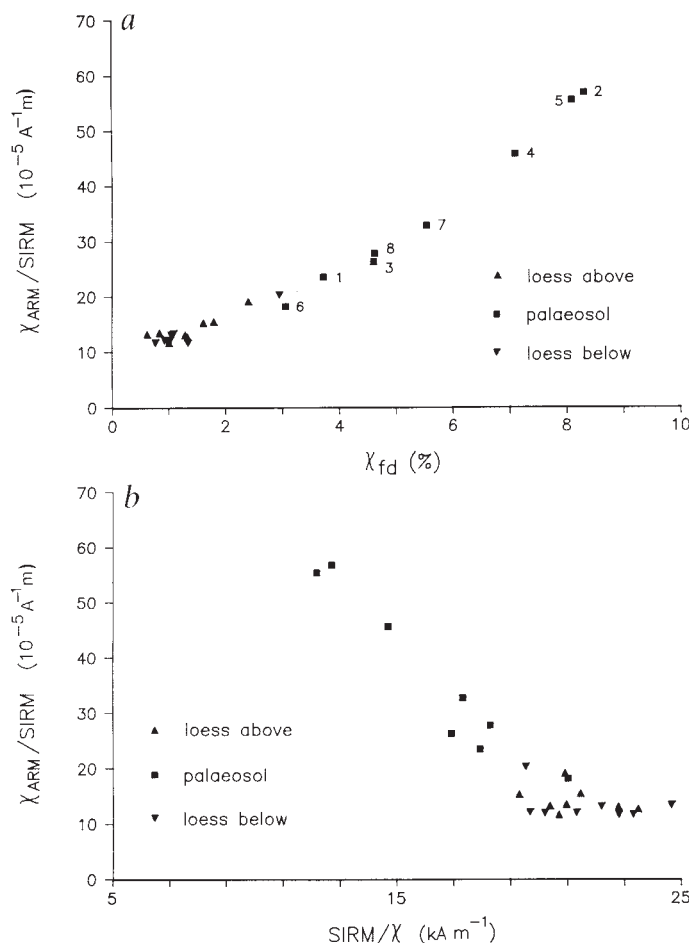


FIG. 2 Plots of a, χ_{ARM}/χ_{IRM} against $\chi_{fd}(\%)$ and b, χ_{ARM}/χ_{IRM} against $SIRM/\chi$ for the Lanzhou samples. A clear relationship between these two normalized properties that differentiate the loess and palaeosol samples in terms of magnetic grain-size proportions is shown in a. The two plots together support the view that $SIRM/\chi$ values in these samples, as in Xifeng section, are depressed in the palaeosols owing to their finer magnetic grain size.

† To whom correspondence should be addressed.

unloading^{2,3}. Using data for the western margin of southern Africa, we demonstrate that upwarps of 600 m with respect to the adjacent continental interior can be generated by this process independently of the mechanics of rifting.

Although the morphology of rifted continental margins is highly variable, two main types of mature rifted margin (>60 Myr old) can be identified: (1) low-lying margins, such as southern Australia, which rise gradually from the coast to a low interior plain (low-elevation rifted margins); (2) steeply rising margins in which the coast is commonly separated from an often relatively high continental interior by a single major escarpment which has retreated inland¹ (high-elevation rifted margins). A typical feature of high-elevation rifted margins, such as those of southwestern Africa, eastern Brazil and eastern Australia, is the gradual decline in mean elevation inland of the escarpment, or series of escarpments (Fig. 1a). This gives rise to a rim of high terrain, or marginal upwarp, typically 300–900 m high and 50–300 km wide which separates the coast from the adjacent continental interior.

Marginal upwarps are also characteristic of young rifted margins, and in this setting they are usually referred to as rift-flank uplifts. Several models have been proposed to explain their development. Thermal effects^{4,5} are time-dependent and decay with a time constant of ~60 Myr (ref. 6) and are therefore not applicable to upwarps associated with mature margins. Other models, including the regional compensation of lithospheric unloading^{7,8} and depth-dependent extension^{9,10} during rifting can generate some residual rift-flank surface uplift, but in each case its amplitude is reduced by sediment loading within the rift. More generally, marginal upwarps along mature rifted

margins cannot be adequately explained by models based on dynamic models of rifting because the axis of maximum surface uplift along most mature rifted margins is now located 100 km or more from the hinge line marking the boundary between rifted and unrifted continental lithosphere. Underplating models¹¹ alone are also incapable of accounting for marginal upwarps because the domal uplifts generated by underplating create broad topographic swells of significantly greater extent than those characteristic of mature rifted margins¹². Our proposed mechanism for the creation of marginal upwarps on mature rifted margins focuses on the isostatic response of the lithosphere to the contrasting rates of denudation that prevail on the coastal and inland flanks of the marked topographic discontinuity present along high-elevation rifted margins. The fact that such a discontinuity is found on both young rifted margins (such as those bordering the Red Sea¹³) and mature rifted margins indicates that it is a persistent feature of margin evolution.

Before rifting, the topography of supercontinents is probably one of predominantly low local relief owing to prolonged denudation, although regional elevation may be relatively high in areas remote from pre-rifting coastlines where local channel gradients, and hence rates of denudation, are low. In addition, underplating associated with rifting may generate a regionally extensive increase in elevation along some rifted margins soon after rifting begins¹². Continental rupture will create significantly lower base levels along newly formed high-elevation rifted margins, and consequently two distinct denudational systems will be established, separated by a retreating major escarpment. This is illustrated by the western margin of southern Africa where

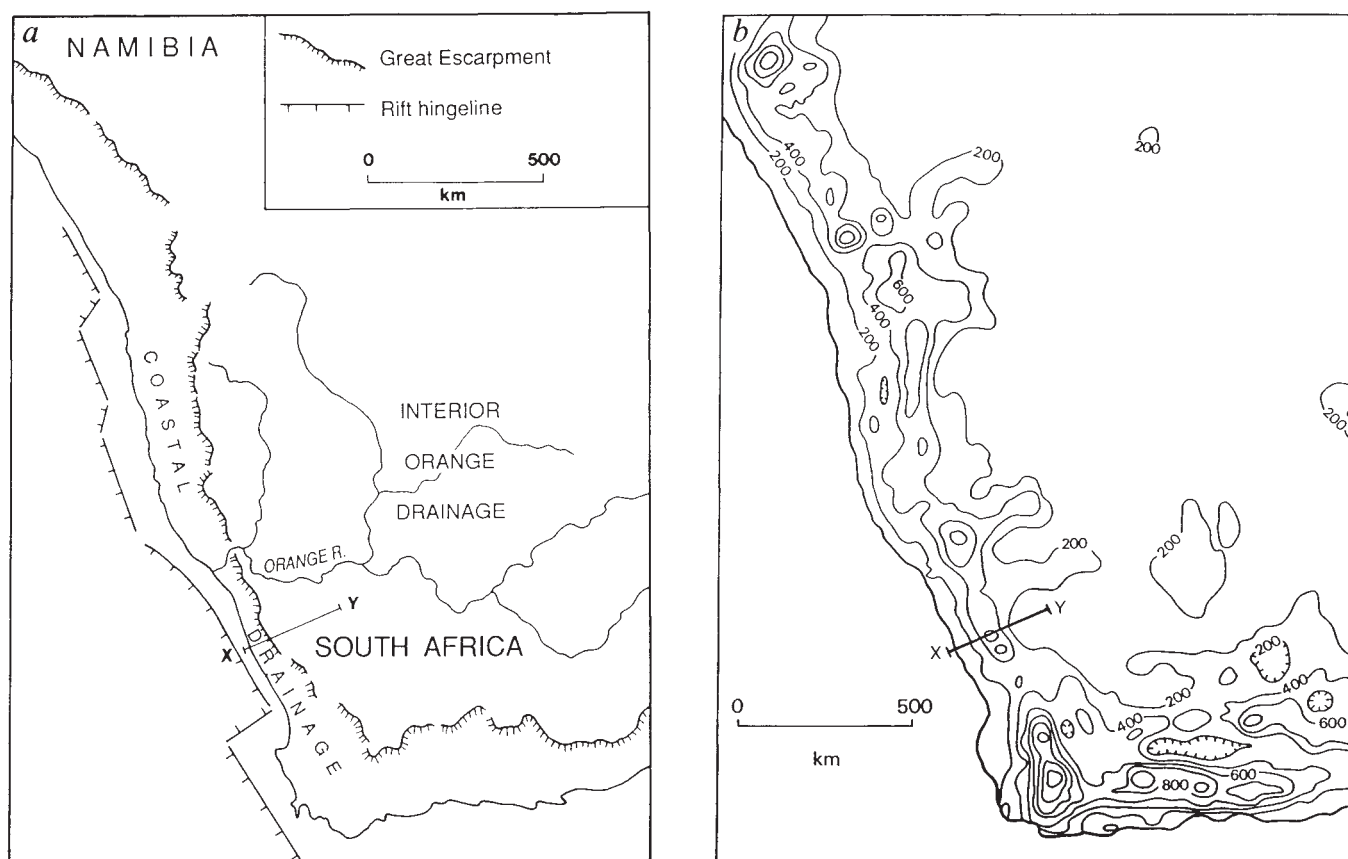


FIG. 1 Major morphological features (a) and variations in local relief (b) along the western margin of southern Africa. Local relief (maximum–minimum elevation) contours (m) have been generalized from local relief

values derived from the 10-min digital topographic data set from the National Geophysical Data Center, Boulder, Colorado. The modelled topographic profile illustrated in Fig. 2 is shown by X–Y line on both maps.

the high local relief along the coast associated with high slope and channel gradients promotes relatively high denudation rates, whereas the very low local relief inland of the Great Escarpment gives rise to low denudation rates, although the regional elevation is relatively high (generally $>1,000$ m) (Fig. 1b). Along this margin the prominence of the escarpment is locally emphasized by the strong lithological control exerted by certain units in the Karoo sequence.

This denudational model is supported by several kinds of evidence. The overall depth of post-rifting denudation averaged over all Atlantic-draining catchments in southern Africa has been estimated to be $\sim 1,800$ m on the basis of offshore sediment volume data¹⁴. Apatite fission-track data indicate upwards of 2.5 km of post-rifting denudation along the margin¹⁵ and therefore denudation rates in the interior must have been significantly lower. Fission-track analyses from other rifted margins also indicate significant post-rifting denudation oceanward of their escarpments^{13,16}. Furthermore, given the former greater westward extent of the Karoo Basin¹⁷, the exposure of older Karoo strata and basement rocks along the western margin of southern Africa also indicates significant depths of erosion, whereas the preservation of younger Karoo sediments and Cenozoic

weathering profiles indicates generally low rates of post-rifting denudation in the interior.

We have modelled the denudation of two terrain types characterized by high and low local relief and separated by a moving boundary. The present mean local relief, based on 10-min areas ($\sim 18 \times 18$ km), of the coastal catchments south of the mouth of the Orange River is 456 m, whereas that of the interior Orange catchment is 191 m. Given the linear relationship between mean local relief and denudation rate¹⁸, this indicates that the present mean denudation rate of the coastal catchments is ~ 2.4 times that of the interior. Given the morphological constraints provided by the present topography and the significantly greater depths of post-rifting denudation along the coast we have assumed that this differential has been maintained since rifting of the Cape Basin (~ 150 Myr ago)¹⁹. We have modelled the Great Escarpment as retreating from the hinge line to its present location at a constant rate of 667 m Myr^{-1} (equivalent to a denudation rate of 900 m Myr^{-1}). The size of the coastal catchments has been assumed to have increased linearly from zero at the time of rifting to a present area of $686 \times 10^3 \text{ km}^2$ as the Great Escarpment retreated, whereas the total Atlantic-draining area has been assumed to have remained constant at $1.6 \times 10^6 \text{ km}^2$. Although there is some evidence that the drainage of the interior of southern Africa was not integrated with Atlantic drainage until some time after rifting^{2,14}, we have adopted the simplest interpretation in view of the uncertainty of the detailed drainage history. In any case the presence of internal drainage in the initial post-rifting phase would have enhanced the differential denudation that we are modelling. Estimating the mean depth of denudation to be 1,800 m for all Atlantic-draining catchments, the mean denudation rates for the coastal and interior Orange catchments are estimated to be 16.5 m Myr^{-1} and 6.9 m Myr^{-1} , respectively. At the coastal boundary the topography is graded to sea level as the escarpment retreats.

In previous attempts to assess the isostatic response to denudational unloading it has either been assumed that denudation rates are linearly proportional to mean regional elevation^{20,21}, or that denudation follows a simple diffusion equation in which it is assumed that flux of regolith is proportional to the slope gradient²². Neither approach provides a valid basis for modelling denudation rates on high-elevation rifted margins. The relationship between regional elevation and denudation rate does not apply because on high-elevation rifted margins low local relief (low slope and channel gradients) prevails on the high interior inland of the escarpment, whereas high local relief typifies the lower elevation terrain on its oceanward flank. Although the diffusion equation may be applicable to homogeneous, regolith-mantled hillslopes²³ it is not capable of modelling regional-scale topography characterized by marked lithological control.

The precise pre-rifting elevation of southern Africa is unknown; nevertheless, taking the mean elevation of present-day Atlantic-draining catchments of ~ 930 m and the depth of post-rifting denudation of 1,800 m, the mean pre-rifting elevation of the region of Atlantic-draining catchments can be estimated by reloading the material represented by offshore sediment thickness (d) and allowing for adjustment by Airy isostasy. The pre-rifting elevation will have been higher than present elevation by h , where $h = d(\rho_m - \rho_c)/\rho_m$. For a crustal density ρ_c of $2,800 \text{ kg m}^{-3}$ and mantle density ρ_m of $3,300 \text{ kg m}^{-3}$ the mean pre-rifting elevation of the Atlantic-draining catchment area is estimated to have been $\sim 1,200$ m. In this simple model we have ignored surface uplift associated with the thermo-mechanical evolution of the margin because it would be substantially compensated by sediment loading offshore. To assess specifically the development of the marginal upwarp, the initial topography is assumed to be a block terminating at the rift hinge line.

The isostatic response to denudational unloading has been modelled assuming a thin elastic plate of constant thickness

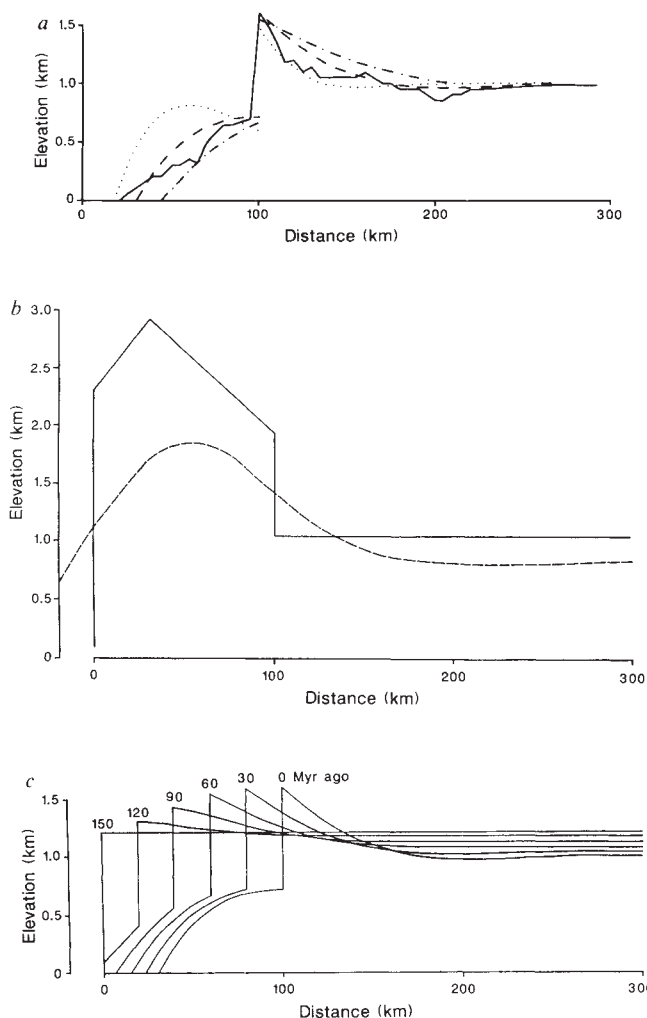


FIG. 2 a, Calculated topographic profiles for various T_e values (dotted line, $T_e = 7.5 \text{ km}$; dashed line, $T_e = 15.0 \text{ km}$; dashed-dotted line, $T_e = 22.5 \text{ km}$) compared with a typical topographic profile (solid line) across the western margin of southern Africa. b, Total denudation (solid line) and total flexural uplift (dashed line) of the margin. c, Evolution of the calculated profiles for $T_e = 15 \text{ km}$. Profile location (X-Y) is indicated on Fig. 1. Zero on horizontal axes indicates location of hinge line. Flexural model parameters were $E = 1 \times 10^{11} \text{ N m}^{-2}$, $\nu = 0.25$, $g = 9.81 \text{ m s}^{-2}$, $\rho_m = 3,300 \text{ kg m}^{-3}$.

overlying an inviscid fluid according to

$$\frac{\partial^2}{\partial x^2} \left(D \frac{\partial^2 w}{\partial x^2} \right) + (\rho_m - \rho_i) g w = l(x)$$

where w is the vertical flexure, D is the flexural rigidity, ρ_m is the mantle density, ρ_i is the flexure infill density (in this case air), g is the acceleration due to gravity, and $l(x)$ is the denudational load as a function of x (ref. 24). Flexural rigidity is given by

$$D = \frac{ET_e^3}{12(1-\nu^2)}$$

where E is Young's modulus, T_e is the effective elastic thickness of the lithosphere and ν is Poisson's ratio. Vertical flexure was calculated in the wavenumber domain from

$$W(k) = R(k) \cdot L(k)$$

where $R(k)$ is the isostatic response function so that

$$R(k) = \frac{1}{(\rho_m - \rho_i)g + Dk^4}$$

$W(k)$ and $L(k)$ are the Fourier transforms of $w(x)$ and $l(x)$ and k is wavenumber given by $k = 2\pi/\lambda$ where λ is the wavelength.

Figure 2a shows topographic profiles calculated using different T_e values compared with a typical profile across the western margin of southern Africa (Fig. 1). The elastic thickness of the highly faulted lithosphere below the Cape Basin has been estimated at 5–10 km (ref. 25) whereas that of the cratonic interior is thought to be 64–90 km (ref. 26). Our best fit of 15 km is therefore consistent with this range of values given the intermediate location of the topographic profile between these two types of flexural terrane. In the region of the escarpment our model generates a maximum of ~600 m of surface uplift with respect to the slowly eroding continental interior and closely replicates the characteristic curved profile of the marginal upwarp. Given the spatial variation in total denudation (Fig. 2b) (which has also been recorded on other rifted margins²⁷), and assuming that compensation to unloading occurs flexurally, the creation of a marginal upwarp is inevitable. Varying the T_e changes the distribution of the upwarped region but the zone of maximum surface uplift remains in approximately the same position. The mismatch between modelled ($T_e = 15$) and actual profile on the inland flank of the upwarp is probably attributable to enhanced rates of denudation in this zone associated with back-cutting tributary drainage systems, which have not been incorporated into the model. The evolution of the marginal upwarp as the escarpment retreats is shown in Fig. 2c.

The generation of marginal upwarps as a result of differential denudational unloading not only explains their presence along mature rifted margins but also has important geomorphological and sedimentological implications. The presence of marginal upwarps up to 100 Myr or more after rifting may account for the deflection of drainage systems away from many mature rifted margins^{2,12,28}. The persistence of such a barrier to sediment transport across high-elevation rifted margins suggests that patterns of offshore deposition are likely to be controlled by a dual drainage system comprising numerous small aggressively eroding coastal catchments and widely spaced outlets of large basins draining continental interiors well into the mature rifted margin stage. Where drainage systems breach a marginal upwarp, the isostatically generated uplift will maintain a significant change in channel gradient (such as the Augrabies Falls on the Orange River). Moreover, the predicted inland migration of the axes of marginal upwarps in response to escarpment retreat has implications for existing denudation and surface uplift chronologies for the Gondwana continents²⁹. Further development of our model requires an assessment of the role of denudational unloading along other rifted margins and the

more detailed incorporation of apatite fission-track data which, in southwestern Africa at least, indicate significant denudation inland of the present escarpment shortly after rifting during the Early Cretaceous¹⁵. □

Received 13 February; accepted 29 June 1990.

- Ollier, C. D. (ed.) *Z. Geomorph. Suppl.* **54** (1985).
- Summerfield, M. A. in *Tectonic Geomorphology* (eds Morisawa, M. & Hack, J. T.) 27–51 (Allen & Unwin, Boston, 1985).
- Thomas, M. F. & Summerfield, M. A. in *Int. Geomorph. 1985, Proc. 1st Int. Conf. on Geomorphology* (eds Gardiner, V. et al.) 935–956 (Wiley, Chichester, 1987).
- Buck, W. R. *Earth planet. Sci. Lett.* **77**, 362–372 (1986).
- Buck, W. R., Martinez, F., Steckler, M. S. & Cochran, J. R. *Tectonics* **7**, 213–234 (1988).
- McKenzie, D. P. *Nature* **307**, 616–618 (1984).
- Braun, J. & Beaumont, C. *Geology* **17**, 760–764 (1989).
- Weissel, J. K. & Karner, G. D. *J. geophys. Res.* **94**, 13919–13950 (1989).
- Royden, L. & Keen, C. E. *Earth planet. Sci. Lett.* **51**, 343–361 (1980).
- White, N. & McKenzie, D. P. *Geology* **16**, 250–253 (1988).
- White, R. S. & McKenzie, D. P. *J. geophys. Res.* **94**, 7685–7729 (1989).
- Cox, K. G. *Nature* **342**, 873–877 (1989).
- Bohannon, R. G., Naeser, C. W., Schmidt, D. L. & Zimmermann, R. A. *J. geophys. Res.* **94**, 1683–1701 (1989).
- Rust, D. J. & Summerfield, M. A. *Mar. Petrol. Geol.* (in the press).
- Brown, R. W., Rust, D. J., Summerfield, M. A., Gleadow, A. J. W. & De Wit, M. C. J. *Nucl. Tracks* (in the press).
- Moore, M. E., Gleadow, A. J. W. & Lovering, J. F. *Earth planet. Sci. Lett.* **78**, 255–270 (1986).
- Oelofsen, B. W. *Geophys. Monogr.* **41**, 131–138 (1987).
- Ahnert, F. *Am. J. Sci.* **268**, 243–263 (1970).
- Gerrard, I. & Smith, G. C. *Am. Ass. Petrol. Geol. Mem.* **34**, 49–74 (1980).
- Stephenson, R. *Geophys. J. R. astr. Soc.* **77**, 385–413 (1984).
- Lambeck, K. & Stephenson, R. *Aust. J. Earth Sci.* **33**, 253–270 (1986).
- Moretti, I. & Turcotte, D. L. *J. Geodyn.* **3**, 155–168 (1985).
- Culling, W. E. H. *J. Geol.* **73**, 230–254 (1965).
- Nadai, A. *Theory of Flow and Fracture of Solids* (McGraw-Hill, New York, 1963).
- Karner, G. D. & Watts, A. B. *J. geophys. Res.* **87**, 2923–2948 (1982).
- Ebinger, C. J., Bechtel, T. D., Forsyth, D. W. & Bowin, C. O. *J. geophys. Res.* **94**, 2883–2901 (1989).
- Speight, J. G. in *The Age of Landforms in Eastern Australia: Tech. Memo. 87/2*, 61–65 (CSIRO Div. Water and Land Resources, Canberra, 1987).
- Summerfield, M. A. *Area* **13**, 3–8 (1981).
- Partidge, T. C. & Maud, R. R. *S. Afr. J. Geol.* **90**, 179–208 (1987).

ACKNOWLEDGEMENTS. We thank N. J. Kusznir and D. J. Rust for discussions. This work was supported by NERC. M.A.S. thanks the Royal Society of Edinburgh for a Support Research Fellowship.

Pyrite formation linked with hydrogen evolution under anaerobic conditions

E. Drobner*, H. Huber*, G. Wächtershäuser†‡§, D. Rose‡ & K. O. Stetter*

* Lehrstuhl für Mikrobiologie, Universität Regensburg, Universitätsstrasse 31, D-8400 Regensburg, FRG

† Tal 29, D-8000 München 2, FRG

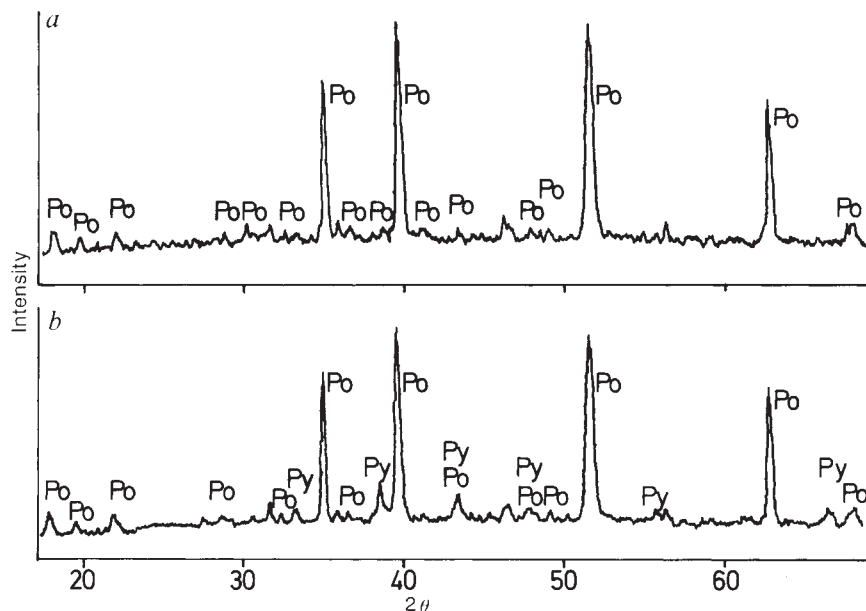
‡ Staatliches Forschungsinstitut für Angewandte Mineralogie, Kumpfmühlerstrasse 2, D-8400 Regensburg, FRG

THE formation of pyrite (FeS₂), an important factor in determining the global redox balance¹, has recently attracted biological interest as a possible direct source of energy for early life^{2–5}. The theory implies that carbon dioxide fixation, in competition with hydrogen formation, can serve as the electron sink for pyrite formation and it seems to be supported by the detection of minute grains of pyrite and iron sulphides inside bacteria^{5–8}. Yet it clashes with the conventional assumption that elemental sulphur or a sulphur equivalent (polysulphide or thiosulphate) is the mandatory oxidant for pyrite formation^{9,10}. It has been stressed that the reaction $\text{FeS} + \text{H}_2\text{S} \rightarrow \text{FeS}_2 + \text{H}_2$ (with H^+ as the oxidant) has “never been observed ... during several years of experimentation”¹⁰. Here we report the formation of both pyrite and molecular hydrogen under fastidiously anaerobic conditions in the aqueous system of FeS and H₂S.

Of the geochemical environments in which pyrite can form, two are of particular biological significance: sedimentary systems, in which pyrrhotite (Fe_{1–x}S) is extremely rare¹¹ and in

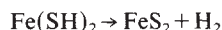
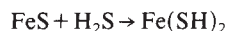
§ To whom correspondence should be addressed.

FIG. 1 X-ray diffraction pattern of pyrrhotite (99%; Johnson Matthey). *a*, Starting material. *b*, After incubation at 100 °C for 14 d in aqueous solution in the presence of H₂S. Po, pyrrhotite; Py, pyrite; 2θ, angle of reflection for Co Kα radiation.



which pyrite seems to be formed from amorphous FeS^{10,12}, and hydrothermal systems in which pyrite may be formed not only from amorphous FeS but also from pyrrhotite¹¹. We have modelled these by reacting aqueous H₂S at 100 °C for 14 days, under strictly anaerobic and nearly neutral conditions, either with amorphous FeS, precipitated from aqueous FeSO₄, or with synthetic (metal basis) pyrrhotite. Our experiments show a linkage between pyrite formation (ascertained by X-ray diffraction) and hydrogen evolution (determined by gas chromatography). Typical results are shown in Table 1 and in Figs 1–3. The pyrrhotite crystals (runs 1, 2) seem to acquire a surface coating of pyrite as indicated by the hollow shells that remain if the pyrrhotite is leached out with HCl (Fig. 2). Amorphous FeS, precipitated in the absence of Fe³⁺ and of sulphur or sulphur equivalents (run 3), seems to produce pyrite in the form of discrete crystals (Fig. 3) as well as some mackinawite (FeS_{1-x}). Hydrogen evolution is also observed with an amorphous FeS precipitated with Na₂S·9H₂O (containing polysulphide) without removing Fe³⁺ (run 4).

The results may best be represented by the following overall reactions



The reaction mechanism may be a concerted four-centre reaction with a simultaneous formation of the covalent bonds of H₂ and S₂²⁻, but radical or hydride involvement cannot be excluded. The reaction of pyrrhotite seems to be a surface redox reaction with a concomitant rearrangement of the S²⁻/S₂²⁻ lattice and a diffusion of ferrous ions. The electron transfer may well be promoted by the semiconductor properties of pyrite.

The considerable variety of geochemical environments that give rise to pyrite may indicate a variety of pyrite-forming pathways using different oxidants. By contrast, the conventional assumption that pyrite formation requires not only H₂S (for FeS formation) but also elemental sulphur or polysulphide as the oxidant placed severe restrictions on the possible geochemical explanations of pyrite formation. In particular, it was difficult to explain sedimentary pyrite formation under anaerobic conditions. For example, Boesen and Postma¹² demonstrated that the freshwater Ancyclus clay in the Gotland deep of the Baltic Sea (deposited 9,200–7,700 years ago) is presently being sulphidized by the downward diffusion of H₂S from the overlying marine muds. In this process the ferrous ions in the clay are first converted into FeS and subsequently into FeS₂. On this conversion they comment "... unless a hitherto unknown oxidant exists the only explanation seems to be either downward diffusion of

TABLE 1 Products of anaerobic FeS–H₂S systems

No.	Starting materials		Products after 14 days	
	FeS	H ₂ S (mmol)	H ₂ (μmol)	Mineral products
1a	pyrrhotite 99%*	2	23 ± 3.5	pyrite
1b	(200 mg)	—	0.25	—
2a	pyrrhotite 99.99%*	2	18	pyrite
2b	(200 mg)	—	0.2	—
3a	FeS amorphous, wet†	2	15 ± 4	pyrite + mackinawite
3b	(precipitated with H ₂ S) (2 mmol)	—	0	—
4a	FeS amorphous, dried‡	2	40 ± 2.5	pyrite + mackinawite
4b	Na ₂ S (200 mg)	—	0.2	—
5	—	2	0.2	—
6	—	—	0	—

All procedures were carried out under CO₂. The solutions were prepared from doubly distilled water, through which N₂–CO₂ had been bubbled for 2 h. Serum bottles (120 ml) were charged with the suspension of FeS, stoppered and supplied with a N₂–CO₂ atmosphere (80:20, 100 kPa) and then charged with an injection of 2 mmol H₂S gas and adjusted to pH 6.5 with NaOH. The H₂S gas was prepared by adding 50% H₂SO₄ to Na₂S·9H₂O in an evacuated serum bottle. During incubation for 14 d at 100 °C in a rotary shaker (100 r.p.m.), the serum bottles were kept in anaerobic cylinders with an N₂–CO₂ atmosphere (80:20, 180 kPa). H₂ was determined by gas chromatography (Hewlett Packard 5890). A packed column filled with Molecular Sieve 5A (Supelco) was used (injection temperature, 190 °C; oven temperature, 140 °C; detection temperature, 220 °C; carrier gas, N₂). For runs 1, 3 and 4, the averages and the standard deviations of the H₂ measurements of three repeats of the reaction are given. Run 2 was not repeated. The traces of H₂ in control runs 1b, 2b and 4b are barely above the background (detection limit 0.1 μmol) and may be due to the reaction 2FeS + 2H⁺ → FeS₂ + Fe²⁺ + H₂. In control run 5, the trace of H₂ may be due to thermal decomposition (H₂S ⇌ H₂ + S). The solid phase was dried in an anaerobic chamber (N₂:H₂ = 95:5) and the mineral composition was analysed by X-ray diffraction.

* Pyrrhotite (99% or 99.99%) (Johnson Matthey) was suspended in 10 ml H₂O. Both pyrrhotites were free of elemental iron as indicated by the lack of hydrogen evolution upon dissolution in concentrated HCl.

† Amorphous wet FeS was precipitated directly in the serum bottles used in the experiment by adding 2 mmol H₂S gas to 10 ml 0.2 M FeSO₄ which had previously been freed of Fe³⁺ by treatment with elemental zinc at 60 °C for 2 h.

‡ Amorphous, dried FeS was prepared in an anaerobic chamber by adding Na₂S·9H₂O (130 g) to 0.6 M FeSO₄ that had not been freed of Fe³⁺ filtering the precipitate, washing it with H₂O and drying it under CO₂. The dried precipitate was suspended in 10 ml H₂O.

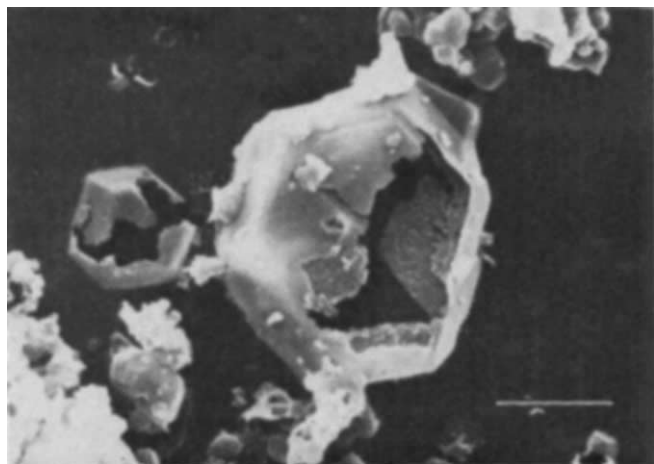


FIG. 2 Scanning electron micrograph of pyrite shells with pyrrhotite pseudomorphy (run 1a) after dissolution of pyrrhotite by treatment with 6N HCl. Scale bar, 0.5 μm .

polysulfides or that the FeS/FeS_2 distribution reflects a historical change [of the redox conditions]. The latter possibility seems, however, unlikely . . .” Our results show that the system $\text{FeS}-\text{H}_2\text{S}$ is a powerful reducing agent and that H^+ can indeed serve as oxidant for pyrite formation. Thus, the somewhat implausible assumption of a slow diffusion of metastable polysulphides in, for example, the sediments of the Gotland deep is no longer required.

The pyritization of fossils is also a highly variable and complex process. For some of these pyritization processes, our results seem to offer straightforward explanations if we assume: (1) diffusion of dissolved ferrous ions and H_2S , at concentrations too small for FeS precipitation (at lower pH, for example) into the site of fossilization; (2) the formation of FeS_2 and H_2 under the pH conditions of the site of fossilization; (3) the diffusion of H_2 out of the site of fossilization. As a variation of this scheme, we may assume diffusion of sulphate ions into the site of fossilization and the formation of hydrogen sulphide by organotrophic sulphate-reducing bacteria at the site of fossilization. In any case, the implausible assumption of a conversion of hydrogen sulphide into elemental sulphur¹³ under the reducing conditions at the site of fossilization is no longer necessary.

The production of H_2 in nature has previously been attributed to biogenic origins, geothermal exhalations and tropospheric

decompositions¹⁴. Our results establish the system $\text{FeS}-\text{H}_2\text{S}$ as an alternative source of hydrogen. This source is ubiquitous, which correlates well with the ubiquitous occurrence of hydrogen-consuming bacteria, such as methanogens, sulphur and sulphate reducers and hydrogen-oxidizing bacteria. Finally, our findings suggest that a functional evolutionary connection might exist between the hydrogen-producing system $\text{FeS}-\text{H}_2\text{S}$ and the hydrogen-producing iron-sulphur centres of hydrogenases and nitrogenases. $\square$

Received 26 April; accepted 10 July 1990.

1. Jorgensen, B. in *Autotrophic Bacteria* (eds Schlegel, H. G. & Bowien, B.) 117–146 (Science Tech, Madison, 1989).
2. Wächtershäuser, G. *Syst. appl. Microbiol.* **10**, 207–210 (1988).
3. Wächtershäuser, G. *Microbiol. Rev.* **52**, 452–484 (1988).
4. Wächtershäuser, G. *Proc. natn. Acad. Sci. U.S.A.* **87**, 200–204 (1990).
5. Popper, K. R. *Nature* **344**, 387 (1990).
6. Farina, M., Esquivel, D. M. S. & Lins de Barros, H. G. P. *Nature* **343**, 256–258 (1990).
7. Mann, S., Sparks, N. H. C., Frankel, R. B., Bazylinski, D. A. & Jannasch, H. W. *Nature* **343**, 258–261 (1990).
8. Williams, R. J. P. *Nature* **343**, 213–214 (1990).
9. Roberts, W. M. B., Walker, A. L. & Buchanan, A. S. *Miner. Deposita* **4**, 18–29 (1969).
10. Berner, R. A. *Am. J. Sci.* **268**, 1–23 (1970).
11. Hall, A. J. *Miner. Mag.* **50**, 223–229 (1986).
12. Boesen, C. & Postma, D. *Am. J. Sci.* **288**, 575–603 (1988).
13. Allison, P. A. in *Palaeobiology* (eds Briggs, D. E. G. & Crowther, P. R.) 253–255 (Blackwell Scientific, Oxford, 1990).
14. Schlegel, H. G. in *Autotrophic Bacteria* (eds Schlegel, H. G. & Bowien, B.) 305–330 (Science Tech, Madison 1989).

ACKNOWLEDGEMENTS. We thank L. Schwarzfischer-Pfeilschifter for technical assistance and K. Popper for improvements of the presentation. This work was supported by the Bundesministerium für Forschung und Technologie and the Fonds der Chemischen Industrie.

Source-area determination of elephant ivory by isotopic analysis

N. J. van der Merwe*†, J. A. Lee-Thorp*,
J. F. Thackeray*, A. Hall-Martin†, F. J. Kruger§,
H. Coetzee§, R. H. V. Bell¶ & M. Lindeque||

* Department of Archaeology, University of Cape Town, Rondebosch 7700, South Africa

† Departments of Anthropology and Earth and Planetary Sciences, Harvard University, Cambridge, Massachusetts 02138, USA

‡ National Parks Board, Private Bag X402, Skukuza 1350, South Africa

§ Bernard Price Institute of Geophysical Research, University of the Witwatersrand, PO Wits, Johannesburg 2050, South Africa

¶ Integrated Rural Development Programme, PO Box 510249, Chipata, Zambia

|| Etosha Ecological Unit, PO Okaukuejo, Outjo, Namibia

RECENT international efforts to conserve the African elephant *Loxodonta africana* prompted us to seek an appropriate method for determining the area from which individual tusks were derived. Trace element analysis of ivory has indicated the potential of chemical analysis for source identification¹, but recent isotopic studies of African mammals^{2–5} suggest another approach. Stable carbon isotope ratios ($^{13}\text{C}/^{12}\text{C}$) in elephant bone collagen clearly reflect the mixture of C_3 foliage and C_4 grasses in the diet, and are directly proportional to the density of C_3 browse². Furthermore, nitrogen isotope ratios ($^{15}\text{N}/^{14}\text{N}$) in bone collagen of African mammals are related to rainfall or water stress^{3–5}. Strontium isotopes ($^{87}\text{Sr}/^{86}\text{Sr}$) in bone or ivory can be expected to reflect local geology^{6,7}. Here we report on a study of ivory and bone samples from different regions of Africa demonstrating the feasibility of tri-variate isotopic analysis to identify the area in which an elephant lived, thus providing a potentially powerful tool for the control of illegal trading in ivory.

For this pilot project, we obtained ivory and/or bone specimens from more than 100 adult elephants, collected by wildlife conservation officers or the police in over 20 identifiable game

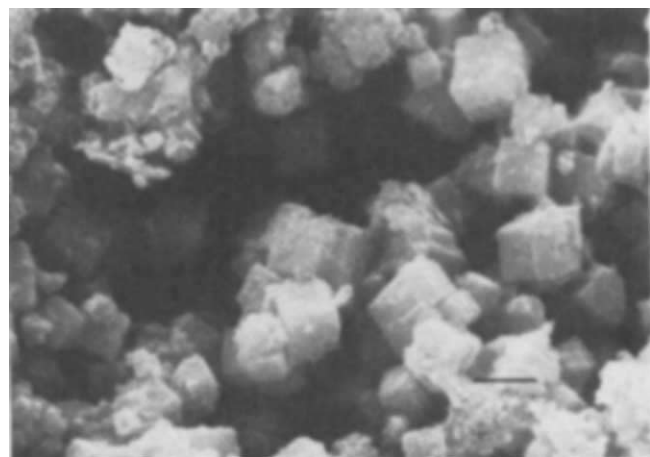


FIG. 3 Scanning electron micrograph of cubic pyrite crystals formed from amorphous FeS (run 4a). Scale bar, 1 μm .

FIG. 1 Localities of African game reserves from which elephant samples were obtained for isotopic analysis.

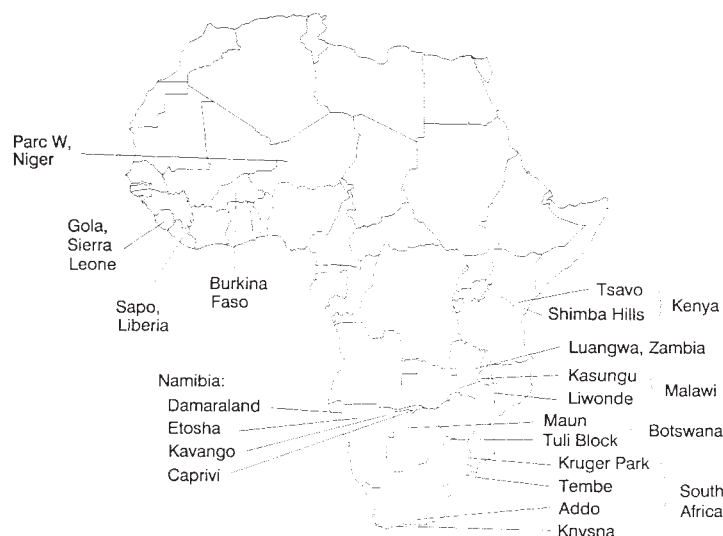


FIG. 2 Distribution of $\delta^{13}\text{C}$ and $\delta^{15}\text{N}$ values obtained from elephant ivory and bone samples from a, West Africa (Liberia, Sierra Leone, Burkina Faso), East Africa (Tsavo, Shimba Hills), and south-central regions (Malawi and Zambia), b, South Africa (Kruger Park, Knysna, Addo), c, Namibia and Botswana. $\delta^{13}\text{C}$ and $\delta^{15}\text{N}$ values serve to distinguish source areas in most cases; strontium isotope ratios (Table 1) serve to distinguish elephants from reserves such as Luangwa and Liwonde that have overlapping $\delta^{13}\text{C}$ and $\delta^{15}\text{N}$ distributions.

refuge areas in 10 African countries (Fig. 1). Collagen was extracted from the bone or dentine samples and converted to carbon dioxide and nitrogen, according to procedures described elsewhere^{2,5,8}. $^{13}\text{C}/^{12}\text{C}$ and $^{15}\text{N}/^{14}\text{N}$ ratios were measured in the VG 602E mass spectrometer of the Archaeometry Laboratory, University of Cape Town; the results are reported as $\delta^{13}\text{C}$ and $\delta^{15}\text{N}$ values, relative to international standards (Table 1). Carbon and nitrogen isotope ratios obtained from ivory and bone are used interchangeably since χ^2 statistical tests indicate no significant difference ($P < 0.05$) between $\delta^{13}\text{C}$ values for samples of dentine and bone from the same individuals (elephants and other African mammals⁸). For $^{87}\text{Sr}/^{86}\text{Sr}$ measurements, strontium nitrate was extracted from ashed samples by standard ion exchange techniques and the isotope ratios determined in the VG 354 thermal ionization mass spectrometer of the Bernard Price Institute for Geophysical Research, University of the Witwatersrand.

$\delta^{13}\text{C}$ and $\delta^{15}\text{N}$ values obtained for elephants from different regions of Africa (Fig. 2) are patterned as predicted. In woodland savannahs, elephants eat primarily C_3 plants (plants carrying out carbon dioxide fixation through the Calvin cycle) and have collagen $\delta^{13}\text{C}$ values near -21.5% (such as Parc W in Niger, Shimba Hills in Kenya, and Skukuza in Kruger Park, South Africa). This value has been well documented for C_3 plant consumers^{9,10}. In dense forests (Sapo, Gola, Knysna), where isotopically light CO_2 is recycled under the forest canopy^{11,13}, $\delta^{13}\text{C}$ values are more negative than -22% ; the denser the forest, the more negative the $\delta^{13}\text{C}$ value. This phenomenon, by itself, tends to distinguish elephants from the forested tropics. Values more positive than -21.5% indicate C_4 grass (plants carrying out carbon dioxide fixation through the C_4 pathway) in the diet as shown, for example, by samples from Tsavo East. An exception to this pattern is Addo, South Africa, where a substantial proportion of the browse consists of CAM plants (those plants in which carbon dioxide fixation proceeds initially through crassulacean acid metabolism) (including *Portulacaria afra*, mean $\delta^{13}\text{C}$ value -17.7% (ref. 2)). In the case of nitrogen isotope ratios, samples from arid zones (northwestern Namibia, Addo and Tsavo East) have high $\delta^{15}\text{N}$ values, as expected for areas where mean annual rainfall is < 400 mm. The CAM plants of Addo are particularly enriched in ^{15}N (ref. 14) and contribute to high $\delta^{15}\text{N}$ values for Addo elephants.

A correlation between $\delta^{13}\text{C}$ and $\delta^{15}\text{N}$ values for elephant samples from areas in southern Africa (Fig. 2) is a reflection of relationships between climate and vegetation. Low $\delta^{15}\text{N}$ values are obtained for samples from areas with relatively high rainfall and a substantial degree of C_3 (bush/tree) vegetation;

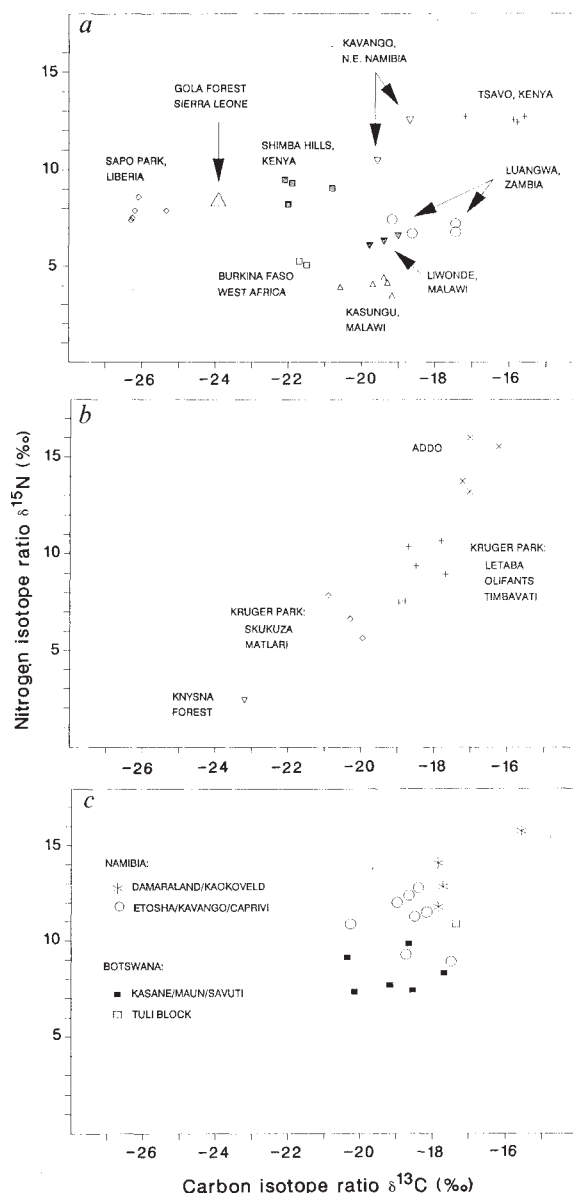


TABLE 1 Isotopic analysis of collagen from elephant bone and dentine

Country	Region	$\delta^{13}\text{C}$	s.d.	n	$\delta^{15}\text{N}$	s.d.	n	$^{87}\text{Sr}/^{86}\text{Sr}$	s.d.	n
Liberia	Sapo	-26.80	0.75	12	7.77	0.46	5	0.71191	—	2
Sierra Leone	Gola forest	-24.00	—	1	8.80	—	1	0.73314	—	1
Burkina Faso		-21.60	—	2	5.20	—	2	0.71493	—	2
Niger	Parc W	-20.83	0.36	5	5.70	—	2	—	—	—
Kenya	Shimba Hills	-21.70	0.52	4	9.04	0.48	4	0.71238	0.00031	3
Kenya	Tsavo	-16.10	0.64	4	12.58	0.13	4	0.71000	—	2
Malawi	Kasungu	-19.62	0.47	6	4.03	0.32	5	0.72683	0.00850	4
Malawi	Liwonde	-19.51	0.32	3	6.40	0.16	3	0.71101	0.00084	3
Zambia	Luangwa Valley	-18.16	0.70	4	6.78	0.52	4	0.72155	0.00168	3
Botswana	(Kasane, Maun, Savuti)	-19.22	0.9	6	8.66	0.68	6	0.72126	0.00071	3
Botswana	Tuli Block	-17.20	—	1	10.80	—	1	0.70649	—	1
Namibia	Damaraland, Kaokoveld	-17.15	1.01	4	13.38	1.44	4	0.72426	0.00068	3
Namibia	Etosha	-18.02	0.33	3	10.55	1.81	3	0.71646	0.00187	5
Namibia	Kavango/Caprivi	-18.85	0.51	5	11.12	1.15	5	0.71928	0.00109	5
South Africa	Kruger Park (east)									
	Olifants	-18.34	—	2	8.25	—	2	0.70799	—	1
	Letaba	-18.60	—	2	9.90	—	2	0.71178	—	2
	Timbavati	-18.25	—	2	8.80	—	2	0.71526	—	1
	Total	-18.40	0.52	6	8.98	1.11	6	0.71171	0.00264	4
South Africa	Kruger Park (west)									
	Matlari	-20.53	0.45	4	7.29	—	2	—	—	—
	Skukuza	-21.16	0.96	7	5.70	—	1	0.72406	—	1
	Total	-20.94	0.87	11	6.77	0.90	3	0.72406	—	1
South Africa	Tembe (Kwazulu)	-19.47	1.14	6	8.88	0.33	4	0.71426	—	2
South Africa	Addo	-17.03	0.37	8	14.64	1.19	6	0.71154	—	2
South Africa	Knysna	-23.20	—	1	2.40	—	1	—	—	—

Mean values and standard deviations obtained for $\delta^{13}\text{C}$, $\delta^{15}\text{N}$ and $^{87}\text{Sr}/^{86}\text{Sr}$ from samples of elephant ivory and bone from African game refuges; most if not all samples were obtained from adult elephants from different herds. Carbon and nitrogen isotope ratios are reported in the δ notation as per thousand (‰) relative to the PDB standard and air N_2 , respectively. Precision (including instrument error) is $<0.1\%$ for $\delta^{13}\text{C}$ and $<0.5\%$ for $\delta^{15}\text{N}$ values. Strontium isotope ratios ($^{87}\text{Sr}/^{86}\text{Sr}$) may be compared with a mean value of 0.710217 ± 0.000030 (precision of 0.004%) for NBS-987 standard.

conversely, high $\delta^{15}\text{N}$ values are obtained for samples from areas with lower rainfall and relatively more C_4 vegetation, notably in the desert regions of Damaraland and Kaokoveld in north-western Namibia.

The numbers of measurements obtained from each refuge in this pilot study do not lend themselves to detailed statistical analysis, but $\delta^{13}\text{C}$ and $\delta^{15}\text{N}$ values are clearly clustered for each region (Fig. 2). For areas where $n \geq 5$, standard deviations are small, between 0.36 and 1.14 for $\delta^{13}\text{C}$, and between 0.32 and 1.19 for $\delta^{15}\text{N}$. Intra-area variation is comparable to intratusk variability, as determined for a tusk from northern Namibia (mean $\delta^{13}\text{C} = -17.6 \pm 0.8\%$, $n = 7$; mean $\delta^{15}\text{N} = 8.1 \pm 0.5\%$, $n = 7$). Intratusk variability is attributable to seasonal variation in the diet and/or seasonal movements, both of which are particularly marked in this desert region. Statistical t -tests indicate significant differences ($P < 0.05$) between mean $\delta^{13}\text{C}$ values for elephant samples from Tsavo East and Shimba Hills in Kenya, only 150 km apart, and between mean $\delta^{15}\text{N}$ values for samples from the same reserves; similarly, mean $\delta^{13}\text{C}$ and $\delta^{15}\text{N}$ ratios for samples from eastern and western sectors of the Kruger Park (Fig. 2b) are significantly different ($P < 0.05$).

Overlap in the distributions of carbon and nitrogen isotope ratios can be expected when dealing with elephants from similar environments. In such cases, strontium isotopes can be used to make distinctions, as $^{87}\text{Sr}/^{86}\text{Sr}$ ratios determined from tusk or bone reflect the age and average Rb/Sr ratios of the parent rocks in a particular region^{6,7}. Young volcanic soils (in the Rift Valley for example) are particularly noticeable in this regard as they have ratios similar to 0.70906, the value for ocean water. Samples from Liwonde (Malawi) and Luangwa (Zambia) have overlapping $\delta^{13}\text{C}$ and $\delta^{15}\text{N}$ distributions (Fig. 2a), but they have clear differences in $^{87}\text{Sr}/^{86}\text{Sr}$ ratios (Table 1). Similarly, $\delta^{13}\text{C}$ and $\delta^{15}\text{N}$ values for samples from Kruger Park (Fig. 2b) overlap with those obtained for elephants from similar environments in

Botswana and northern Namibia (Fig. 2c), but strontium isotope ratios differ (Table 1) because of geological history. In general, samples from areas with soils which developed on old granitic rocks (shield areas) have $^{87}\text{Sr}/^{86}\text{Sr}$ ratios that are >0.715 , whereas those from areas with rocks of basaltic composition have lower ratios (for example, Tuli Block in Botswana).

We conclude that trivariate isotopic analysis provides a potential means of identifying source areas of ivory on a wide geographic scale. Practical application of this method will require a more comprehensive index of ivory (or bone) isotope signatures for all regions where elephants still exist in Africa. Such a database could provide the foundation for international control of the ivory trade and hence for the conservation of the elephant and its habitat. □

Received 26 January; accepted 11 June 1990.

1. Bell, R. H. V., Kelsall, J. P., Rawluk, M. & Avery, D. H. *Pachyderm* **12**, 29–31 (1989).
2. Van der Merwe, N. J., Lee-Thorp, J. A. & Bell, R. H. V. *Afr. J. Ecol.* **26**, 163–172 (1988).
3. Heaton, T. H. E., Vogel, J. C., von la Chevallerie, G. & Collett, G. *Nature* **322**, 823–824 (1986).
4. Ambrose, S. H. & DeNiro, M. J. *Oecologia* **69**, 395–406 (1986).
5. Sealy, J. C., van der Merwe, N. J., Lee-Thorp, J. A. & Lanham, J. L. *Geochim. cosmochim. Acta* **51**, 2707–2717 (1987).
6. Ericson, J. E. *J. hum. Ecol.* **14**, 503–514 (1985).
7. Sealy, J. C., van der Merwe, N. J., Sillen, A., Kruger, F. J. & Krueger, H. W. *J. archaeol. Sci.* (in the press).
8. Lee-Thorp, J. A. PhD thesis, University of Cape Town (1989).
9. Vogel, J. C. S. *Afr. J. Sci.* **74**, 298–301 (1978).
10. Van der Merwe, N. J. & Vogel, J. C. *Nature* **276**, 815–816 (1978).
11. Vogel, J. C. *Oecol. Pl.* **13**, 89–94 (1978).
12. Medina, E. & Minchen, P. *Oecologia* **45**, 377–378 (1980).
13. Van der Merwe, N. J. & Medina, E. *Geochim. cosmochim. Acta* **53**, 1091–1094 (1989).
14. Heaton, T. H. E. *Oecologia* **74**, 236–246 (1987).
15. Sealy, J. C. *Cambridge Monographs in African Archaeology* **15**, British Archaeological Reports International Series 293 (British Archaeological Reports, Oxford, 1986).

ACKNOWLEDGEMENTS. We thank the wildlife conservation officers who provided samples for analysis and C. B. Smith (Bernard Price Institute of Geophysical Research) for cooperation with the strontium measurements. This research was supported by the Foundation for Research Development (South Africa), the South African Nature Foundation, the National Parks Board and the Rhino and Elephant Fund, D. Sholto-Douglas (University of Cape Town) assisted with the analyses. J. Lanham, J. Sealy, A. Sillen and J. C. Vogel commented on the manuscript.

Isotope fingerprints in elephant bone and ivory

J. C. Vogel, B. Eglington & J. M. Auret

EMATEK Division, CSIR, Box 395, Pretoria, South Africa

THE isotopic composition of carbon and nitrogen as well as of strontium in animal bone is related to the environment in which the animal lived¹⁻⁶. It can be assumed that this is also the case for lead isotopes. In theory, therefore, we have a way of pinpointing the origin of elephant ivory, which may be of value in conservation. Here we report that by analysing the isotope ratios of these elements, a clear distinction between several different populations of the African elephant can be made.

We used the isotope signatures to characterize elephant populations in different habitats in southern Africa (Fig. 1). In most cases bone from the jaws or ribs was used, but some samples of ivory were also included. The specimens came from the Knysna Forest on the south coast of the Cape Province, the Addo Elephant National Park in the southwest Cape, different areas of the Kruger National Park, East Transvaal, and the northern Namib Desert in Namibia. Three single samples from areas to the east of the latter group, namely East Kaokoveld, the Etosha Game Reserve and Caprivi, complete the list.

The factors expected to determine the isotope ratios of the various elements in the skeletons of elephants are briefly discussed below. In most parts of Africa the $^{13}\text{C}/^{12}\text{C}$ ratio in animal tissue is determined primarily by the amount of (C_4) grass ingested by the animal¹ and, thus, in the case of mixed feeders such as elephants, which both graze and browse, is to a degree dependent on the environment². The $^{15}\text{N}/^{14}\text{N}$ ratio in bone collagen of ungulates, is, again, higher in more arid habitats^{3,4}. This may be ascribed, at least in part, to the higher ^{15}N levels of the local vegetation⁵, and for the rest it has been proposed to be the effect of drought stress on the protein metabolism of the animals themselves⁶. The carbon and nitrogen isotopes, even when taken together, do not, however, completely characterize the various elephant populations of southern Africa⁷. Only by adding a further independent isotope signature can this be achieved.

The $^{87}\text{Sr}/^{86}\text{Sr}$ ratio in animal bone is determined by that of the strontium in the animal's plant diet, which is determined by the ratios in the soil and water of the area^{8,9}. In most cases these ratios reflect those of the underlying rocks. ^{87}Sr is produced by the slow radioactive decay of ^{87}Rb , so that its abundance depends on the rubidium content and the age of the rock suites in the substrata of the area. Consequently, the higher the Rb/Sr ratio and the older the rock, the higher the $^{87}\text{Sr}/^{86}\text{Sr}$ ratio will be.

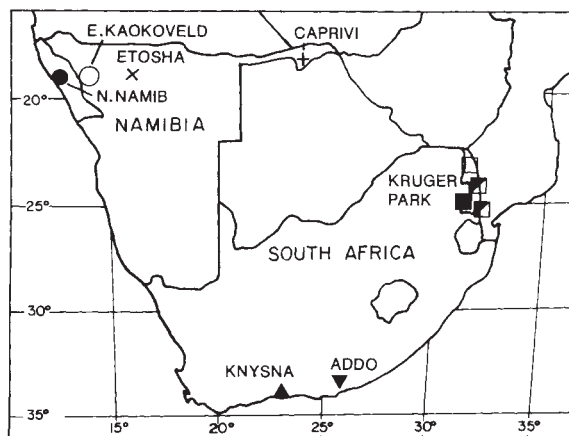


FIG. 1 Map of southern Africa indicating the localities from which samples of elephant bone were analysed. The same symbols are used in Fig. 2.

The lead isotopes, which are decay products of uranium and thorium, have not been used previously on bone material in the same way as have the strontium isotopes. Lead substitutes for calcium in a similar manner to strontium and the lead isotope ratios should thus also be characteristic of the underlying rocks.

Standard procedures were used to prepare samples for isotope analysis in suitable mass spectrometers. The results are listed in Table 1. ^{13}C analysis only distinguishes the Knysna specimens from the rest, whereas ^{15}N clearly distinguishes Knysna, Kruger Park and Addo samples, with the Namibian specimens falling between and overlapping the latter two groups. The ^{87}Sr , again, distinguishes the Namibian samples from the rest, the higher ^{87}Sr values being associated with areas on older, more felsic substrata (see Table 1). It is also noteworthy that the specimens from the Archaean granitic area in the far southwest of the Kruger Park have distinctly higher ^{87}Sr values than do those from the younger Mesozoic basalts in the far northeast, as is to be expected. The three lead isotopes show close correlation with each other, but do not effectively distinguish the different regions from each other, although there is a tendency for the isotope ratios in Addo and most of the Kruger Park specimens to be lower than those in Knysna and the Namib.

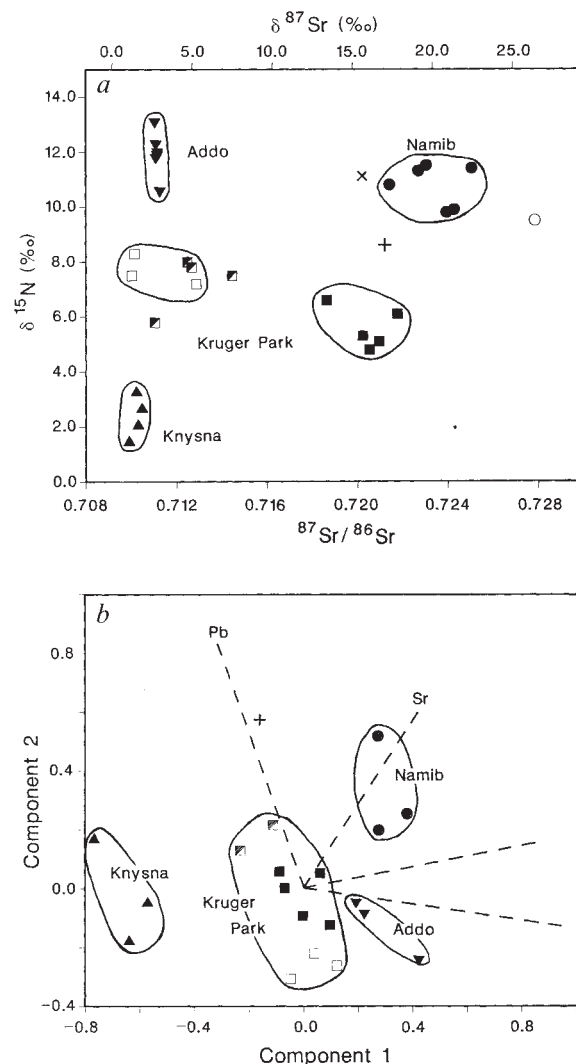


FIG. 2 a, Plot of the relative ^{15}N and ^{87}Sr contents of elephant bone, showing the complete separation of the different elephant populations from Knysna, Addo, Kruger Park and the N. Namib Desert (Table 1). The specimens belonging to specific populations are encircled. b, Plot of the first two principal components of the multivariate statistical analysis of the specimens for which measurements of the isotopic composition of carbon, nitrogen, strontium and lead were made.

TABLE 1 Isotope ratio measurements of elephant bone and tusk from localities in southern Africa*

Sample no.	Locality	Main rock substrate	Material	$\delta^{13}\text{C}$ (‰)	$\delta^{15}\text{N}$ (‰)	$^{87}\text{Sr}/^{86}\text{Sr}$	$\delta^{87}\text{Sr}$ (‰)	$^{206}\text{Pb}/^{204}\text{Pb}$	$^{207}\text{Pb}/^{204}\text{Pb}$	$^{208}\text{Pb}/^{204}\text{Pb}$
B730	Knysna Forest, S Cape Prov	Palaeozoic sandstones, Table Mountain Group	Bone	-23.6	+3.0	0.710609	+2.0	17.295	15.561	37.435
B765			Bone	-23.8	+1.2	0.710281	+1.5	16.919	15.509	36.851
C4099			Tusk	-24.5	+2.4	0.710845	+2.3			
C4100			Bone	-24.7	+1.8	0.710676	+2.1	18.039	15.724	38.905
			Average	-24.2	+2.1	0.710603	+2.0	17.417	15.598	37.730
			s.d.	± 0.5	± 0.8	± 0.000236	± 0.3	± 0.569	± 0.112	± 1.058
B733	Addo Elephant Park, SE Cape Prov.	Mesozoic marine seds, Uitenhage Group	Bone	-19.2	+12.0	0.711482	+3.2			
B734			Bone	-18.8	+11.7	0.711562	+3.3			
B735			Bone	-17.7	+10.3	0.711656	+3.5	17.078	15.502	36.859
B738			Bone	-17.1	+12.8	0.711442	+3.2	16.476	15.458	36.259
B739			Bone	-18.8	+11.6	0.711546	+3.3	17.118	15.538	37.165
B741			Bone	-17.5	+11.5	0.711484	+3.2			
			Average	-18.2	+11.7	0.711529	+3.3	16.891	15.499	36.761
			s.d.	± 0.9	± 0.8	± 0.000077	± 0.1	± 0.360	± 0.040	± 0.461
Kruger Park, E. Tvl*										
B770.1	Far SW. (Skukuza to Malelane)	Archaean granitoid gneisses	Bone	-18.9	+6.3	0.719014	+13.8	16.430	15.440	36.131
B770.3			Bone	-19.2	+5.8	0.720617	+16.1	16.804	15.698	36.821
B770.4			Bone	-20.1	+4.5	0.720915	+16.5	16.932	15.649	36.904
B770.5			Bone	-19.8	+4.8	0.721340	+17.1	16.363	15.419	36.037
B770.0			Bone	-20.3	+5.0	0.722530	+18.8	16.725	15.448	36.482
			Average	-19.7	+5.3	0.720883	+16.5	16.651	15.531	36.475
			s.d.	± 0.6	± 0.7	± 0.001273	± 1.8	± 0.244	± 0.132	± 0.392
B773.1	Far NE (S of Pafuri)	Mesozoic basalts, Karoo Seq.	Bone	-20.0	+7.2	0.710417	+1.7	16.396	15.460	36.130
B773.2			Bone	-18.0	+8.0	0.710528	+1.9	16.612	15.463	36.336
B773.4			Bone	-19.0	+6.9	0.713263	+5.7	16.526	15.478	36.259
			Average	-19.0	+7.4	0.711403	+3.1	16.511	15.467	36.242
			s.d.	± 1.0	± 0.7	± 0.001612	± 2.3	± 0.109	± 0.010	± 0.104
C4069	Far SE (Crocodile Bridge)	Transitional	Tusk	-18.6	+7.2	0.714848	+8.0			
C4068	Mid S	Transitional	Bone	-20.0	+5.5	0.711428	+3.1	17.924	15.647	38.552
C4050	(Tshokwane)		Tusk	-17.2	+7.7	0.712864	+5.2			
C4051			Bone	-19.7	+7.5	0.713054	+5.4	18.040	15.728	38.911
Namibia										
B846	N. Namib Desert	Mid-Late Proterozoic gneisses and schists	Bone	-18.6	+9.5	0.724338	+21.3	17.042	15.529	37.139
B847			Bone	-18.7	+11.1	0.725462	+22.9	18.012	15.718	38.869
B848			Bone	-17.8	+11.2	0.723448	+20.1	17.280	15.548	37.215
B849			Bone	-18.0	+9.6	0.724681	+21.8			
B851			Bone	-18.7	+10.5	0.721813	+17.8			
B853			Bone	-17.9	+11.0	0.723087	+19.6			
			Average	-18.3	+10.5	0.723805	+20.6	17.445	15.598	37.741
			s.d.	± 0.4	± 0.8	± 0.001297	± 1.8	± 0.506	± 0.104	± 0.978
B857	E Kaokoveld	Mid Proterozoic granitoid gneisses	Bone	-16.4	+9.2	0.728249	+26.9			
B316	Etosha Game Reserve	Cenozoic sands, Kalahari Group over	Bone	-17.9	+10.8	0.720615	+16.1			
B112	Caprivi	Late Proterozoic meta-seds	Bone	-21.7	+8.2	0.721590	+17.5	18.485	15.655	38.298

* $\delta^{13}\text{C}$, Carbon isotope ratios expressed relative to the PDB reference standard; $\delta^{15}\text{N}$, nitrogen isotope ratios relative to atmospheric nitrogen; $\delta^{87}\text{Sr}$, strontium isotope ratios relative to average modern-day sea water (0.70920). The reproducibility of the measurements (1 s.d.) is better than 0.4‰ for carbon and nitrogen, 0.1‰ for strontium, and 0.1‰, 0.1‰ and 0.2‰ for the three lead isotope ratios, respectively. Replicates of NBS SRM 987 strontium carbonate gave an isotope ratio of 0.710275 ± 16 (1 s.d., $n=10$). Mass fractionation corrections for the lead ratios are based on replicates of NBS SRM 981.

† The four groups represent four different populations in the Park. The specimens from the Far SW are from at least three different herds; those from the Far NE are from two separate herds.

For the various elephant populations considered here, complete distinction, without any overlap, is achieved by plotting ^{15}N against ^{87}Sr (Fig. 2a).

As further populations from other parts of Africa are added, the characterization based on these two isotopes alone is bound to become less distinct, and multivariate analysis, using all the available isotopes, would be more appropriate. To illustrate this approach, isotope data for the 20 specimens for which lead isotope measurements are also available (using the $^{206}\text{Pb}/^{204}\text{Pb}$ ratio as representative of the three lead isotopes) have been analysed using multivariate statistics, in the form of combined R- and Q-mode components analysis¹⁰, based on a correlation similarity matrix. Ninety-eight per cent of the total variance in the data set is explained by three components which are essentially dominated by the isotope ratios of C and N (component 1), Pb and Sr (component 2) and Sr (component 3). Distinction of the four populations is quite clear on a plot of component 1 against component 2 (Fig. 2b), although plots of the other components could be more effective in distinguishing separate localities within a region.

Most of our analyses were of bone, but there is no reason why ivory cannot also be used. The only difference would be

that a sample of ivory taken from the base of the tusk reflects the food intake during the last year or so of the life of the animal, whereas the bone is integrated over most of its lifetime. To check the validity of using ivory, three pairs of bone and ivory from the same animal were analysed (see Table 1). In all three cases the ivory values fall into the same grouping in Fig. 2a as do those of the bone. It can, therefore, be concluded with confidence that the combined analysis of ^{15}N and ^{87}Sr in tusks can provide a useful tool for establishing their origin. More distinctive fingerprinting can be achieved by adding other isotope signatures, notably those of lead, carbon and, possibly, sulphur. There would have to be a more complete survey of elephant habitats in the rest of Africa before the technique can be applied on a wide scale. In principle, however these isotope fingerprints in ivory can make a positive contribution towards monitoring the products of the endangered African elephant. □

Received 21 March; accepted 11 June 1990.

1. Vogel, J. C. S. *Afr. J. Sci.* **74**, 298-301 (1978).
2. Van der Merwe, N. J., Lee Thorp, J. A. & Bell, R. H. V. *Afr. J. Ecol.* **26**, 163-172 (1988).
3. Heaton, T. H. E., Vogel, J. C., von la Chevallerie, G. & Collett, G. *Nature* **322**, 822-823 (1986).

4. Sealy, J. C., Van der Merwe, N. J., Lee Thorp, J. A. & Lanham, J. L. *Geochim. cosmochim. Acta* **51**, 2707–2717 (1987).
5. Heaton, T. H. E. *Oecologia* **74**, 236–246 (1987).
6. Ambrose, S. H. & DeNiro, M. J. *Oecologia* **69**, 395–406 (1987).
7. Vogel, J. C., Talma, A. S., Hall-Martin, A. & Viljoen, P. J. S. *Afr. J. Sci.* **86**, 147–150 (1990).
8. Ericson, J. E. *J. hum. Evol.* **14**, 503–514 (1985).
9. Sealy, J. C., Van der Merwe, N. J., Sillen, A., Kruger, F. J. & Krueger, H. W. *J. archaeol. Sci.* (in the press).
10. Davis, J. C. *Statistics and Data Analysis in Geology* (Wiley, New York, 1986).

ACKNOWLEDGEMENTS. We thank Dr A. Hall-Martin, Skukuza, and P. J. Viljoen, Pretoria, for supplying most of the samples, G. von la Chevallerie, Y. Stander and G. Jakob for assistance with the laboratory preparations, and S. Talma and T. Heaton for supervising the carbon and nitrogen mass spectrometric analyses.

Maintenance of B-cell memory by long-lived cells generated from proliferating precursors

Birgit Schitteck & Klaus Rajewsky

Institute for Genetics, University of Cologne, Weyertal 121, D-5000 Cologne 41, FRG

A BASIC feature of T-cell dependent antibody responses is the generation of memory: on a second contact with an antigen a secondary response is produced in which somatically mutated antibodies with increased affinity are synthesized¹. Memory can persist for long periods of time. This has classically been ascribed to the generation of long-lived memory B cells^{2,3}. However, it is also possible that persisting antigen, on which memory may depend⁴, maintains a population of cycling memory cells under continuous selection⁵ or continuously recruits newly generated B cells into the memory B-cell compartment⁶. To discriminate between these mechanisms we have now directly analysed the proliferative activity in the memory B-cell compartment of the mouse by measuring bromodeoxyuridine incorporation *in vivo*. We show that after an initial phase of extensive proliferation after primary immunization, memory cells can persist in the organism for extended periods of time in the absence of cell division.

In accordance with the work of Hayakawa *et al.*⁷, we raised specific memory B cells in mice by immunization with the highly fluorescent protein phycoerythrin (PE) and characterized them as PE-binding cells that express the B-cell marker B220 and low levels of surface immunoglobulin differing in class from IgM and IgD as expressed by naive B cells. Using magnetic cell sorting and flow cytometry in combination, we analysed these cells directly for their proliferative activity *in vivo*. The outline of the experiment is given in Fig. 1. At various times after immunization, bromodeoxyuridine (BrdU; 1 mg ml⁻¹) was added to the drinking water of the animals, for various periods of time. Dividing lymphoid cells incorporate BrdU into their DNA under these conditions, and the proliferative behaviour of BrdU-labelled B lineage cells is not different from that of unlabelled cells^{8,9}. At the time of analysis, spleen cells were isolated, enriched for IgM⁺IgD⁺ B cells on a magnetic cell sorter (MACS)¹⁰, and stained with PE and an anti-B220 antibody. The cells could then be processed in either of two ways: flow cytometric analysis and cell sorting allowed quantitation and functional and morphological analysis of PE binding memory cells; alternatively, the cells were fixed, stained with an anti-BrdU antibody and analysed for BrdU incorporation by three-colour flow cytometry.

PE-binding, B220-positive (B220⁺PE⁺) cells were readily detectable among the IgM⁺IgD⁺ spleen cells of PE-immunized mice (see Fig. 2a). When these cells were examined under the fluorescence microscope, they turned out to be small lymphocytes, negative for IgM and IgG1 in the cytoplasm. To confirm that these cells are functionally active memory cells, we tested them for their ability to produce a secondary antibody response *in vitro*, as described earlier⁷. For this purpose, PE-

binding (B220⁺PE⁺) and nonbinding B cells (B220⁺PE⁻) isolated from the spleens of mice immunized with PE 7 or 12 weeks earlier were cultured in the presence or absence of helper T cells (CD4⁺) from PE-immunized mice (Table 1). After 14 days the culture fluids were analysed for the presence of PE-specific IgG1 antibodies. As few as 250 B220⁺PE⁺ cells, but not B220⁺PE⁻ cells produce a strong IgG1 anti-PE-response when (and only when) cultured together with PE-specific helper T cells (Table 1). This confirms that in the PE-primed animals PE-specific memory B cells are located in the B220⁺PE⁺ population and that the antibody response measured is not due to plasma cell contamination.

To determine the proliferative activity of these memory cells, the cells were analysed for BrdU incorporation *in vivo*. Feeding of BrdU to the mice was started either 1 day, 4 weeks, 10 weeks or 20 weeks after priming, and stopped 1, 3, 5 or 7 weeks later; the spleen cells of the animals were processed as shown in Fig. 1. Representative staining patterns are shown in Fig. 2. Plotted is either PE against anti-B220 (Fig. 2a) or PE against anti-BrdU-staining (Fig. 2b). To determine the ratio of labelled versus unlabelled cells among the memory cells, we gated on B220⁺PE⁺ cells (Fig. 2a), collected 1,000–5,000 cells in this fraction, analysed the BrdU-staining profile (Fig. 2c–f) and calculated the percentage of BrdU-labelled cells in this population.

Almost all of the B220⁺PE⁺ cells (92%) divide within the first five weeks after priming (Fig. 3a). That the values do not reach 100% could indicate that this population contains cells that do not divide at all. More likely, this represents a problem in the evaluation of the staining data, as negative and positive populations were not always fully separated. When BrdU-treatment is

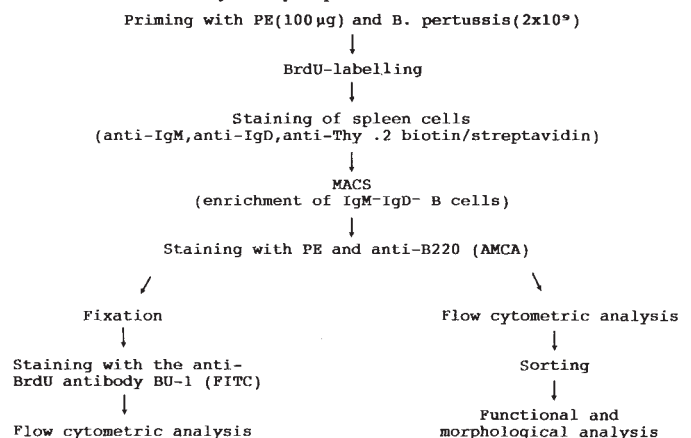


FIG. 1 Schematic outline of the experiment.

METHODS. BALB/c mice or BALB/c mice congenic for the IgH^b haplotype (CB.20) were immunized by intraperitoneal injection of 100 µg PE (gift of Dr W. Müller) and 2 × 10⁹ *Bordetella pertussis* organisms as an adjuvant. No difference in the response of the two strains was observed. BrdU-labelling was done by feeding the mice for various time periods after priming with 1 mg BrdU (Sigma) per ml drinking water^{8,9}. Spleen cells of three mice from a given experimental group were pooled and erythrocytes lysed in 0.8% NH₄Cl. To enrich for IgM⁺IgD⁺ B cells we used the method of magnetic cell sorting as described previously¹⁰. Briefly, spleen cells were stained with the biotinylated antibodies RS3.1 (ref. 19) or MB86 (ref. 20) (anti-IgM), 10.4.22 (ref. 21) or 4/4D7 (a gift from Dr T. Tokuhisa) (anti-IgD) and H013-14 (anti-Thy-1.2)²². The cells were then incubated with streptavidin (Boehringer) and, in a third step, with superparamagnetic biotinylated microparticles. Eight to eighteen percent of the spleen cell population remained unlabelled by this procedure and were enriched to 85–95% using a MACS¹⁰. The recovery was 70–80%. The enriched cells were stained with PE (10 µg per 5 × 10⁷ cells) and with the anti-B220 antibody RA3.6B2 (ref. 23) coupled to AMCA (Amino-methyl-coumarin-acetic acid)²⁴ (10 µg per 5 × 10⁷ cells). Cells were fixed in 70% ethanol, washed in PBS and stained with the anti-BrdU antibody BU-1 (refs 25, 26) (culture supernatant; 30 µg per 2 × 10⁷ cells) (kindly provided by Dr N. Gonchoroff) and with an anti-IgG2a⁹ antibody (Ig(1a)B.3)²¹ coupled to fluorescein-isothiocyanate (FITC; 10 µg per 5 × 10⁷ cells). Using this anti-BrdU antibody it is not necessary to denature the cellular DNA by acid or base treatment, which would destroy PE-staining of the cells.

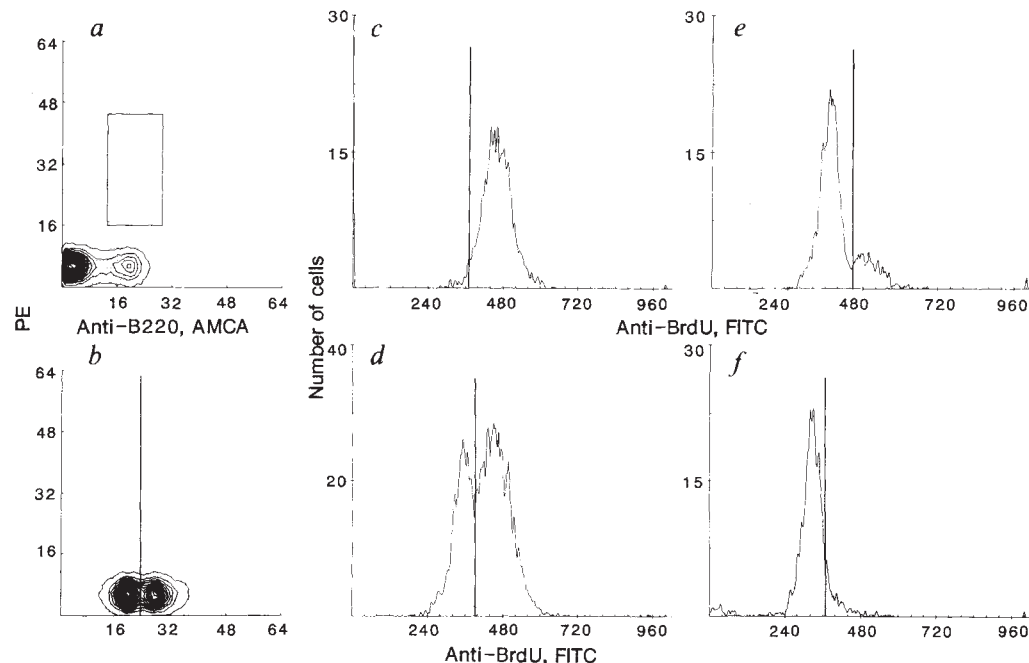
FIG. 2 BrdU-analysis of memory B cells. Shown are representative staining patterns for B220 surface expression and PE-binding (a), or BrdU-incorporation and PE-binding (b) of spleen cells of PE-immunized mice after enrichment of IgM⁺IgD⁺ B cells. After gating on B220⁺PE⁺ cells (see gate in Fig. 2a), between 1,000 and 5,000 of these cells were collected and the BrdU-staining pattern analysed (c-f). Feeding of BrdU was started either 1 day (b, c), 4 weeks (d), 10 weeks (e) or 20 weeks (f) after priming. Feeding was stopped after either 19 days (f), 34 days (b, c and e) or 48 days (d) of BrdU-labelling. Gates were set between positive and negative populations and the percentage of BrdU-labelled versus unlabelled cells calculated according to this gate (see Fig. 3a).

METHODS. Enrichment and staining of spleen cells was done as described in Fig. 1. Fluorescence staining was analysed on a FACS Star Plus using three-colour immunofluorescence detection. List mode data were stored in a MicroVax and analysed using Consort 40, version 5.2 (Becton Dickinson). Only cells falling within

started four weeks after priming and continued for 18 or 33 days, 36% or 62% of the memory B cells are labelled, respectively. Prolongation of the BrdU-treatment to 48 days does not result in a significantly higher percentage of labelled cells in the latter population (64%; Fig. 3a). In contrast, when BrdU-treatment is started 10 weeks after immunization, 82–88% of the memory B cells do not divide within 5 or 7 weeks. A similar situation is encountered 20–23 weeks after priming. The low degree of cell proliferation 10–20 weeks after priming indicates either some recruitment of naive cells into the memory B-cell compartment, or proliferation of some memory B-cell clones still at that time. However, the data clearly show that 9–10 weeks after immunization and later, most of the memory cells do not divide and are not recruited from newly generated B cells over extended periods of time (≥ 7 weeks). The number of B220⁺PE⁺ cells in the spleens of the immunized animals—of the order of 0.01% of the total spleen cell population, a little below the 0.02–0.05% reported by Hayakawa *et al.*⁷—remain essentially constant for at least 200 days, from around 12 days after priming (Fig. 3b). This accords with the observation (K. Hayakawa, personal communication) that functionally active, PE-specific memory B cells persist in the animals for at least 17 months after priming. We thus conclude that after an initial period of extensive proliferation and/or recruitment from newly generated cells, memory B cells convert into a population of long-lived cells.

We consider it likely that the incorporation of BrdU seen in the memory compartment in the first weeks after immunization mainly reflects antigen-driven proliferation of precursor B cells, from which somatic mutants expressing antibodies with high affinity for the antigen are selected¹. It has been suggested^{11,12} that this process of proliferation and selection takes place in the germinal centres, which form after primary, T cell-dependent immunization in the peripheral lymphoid organs. The duration of this early, T cell-dependent^{13,14} phase of memory generation may depend on the nature of the primary antigenic stimulus. The present data indicate that memory cells generated in this process are long-lived cells that can persist in the organism for a good part of its lifetime. This is in accord with recent data demonstrating that the persistence of memory B cells requires little, if any, T-cell help¹⁴.

These experiments do not address the question of whether



normal lymphocyte values for both forward and orthogonal scatter are included⁹. Cells with more than 16–25 units of PE-staining intensity were considered as PE-bright (PE⁺) (see a).

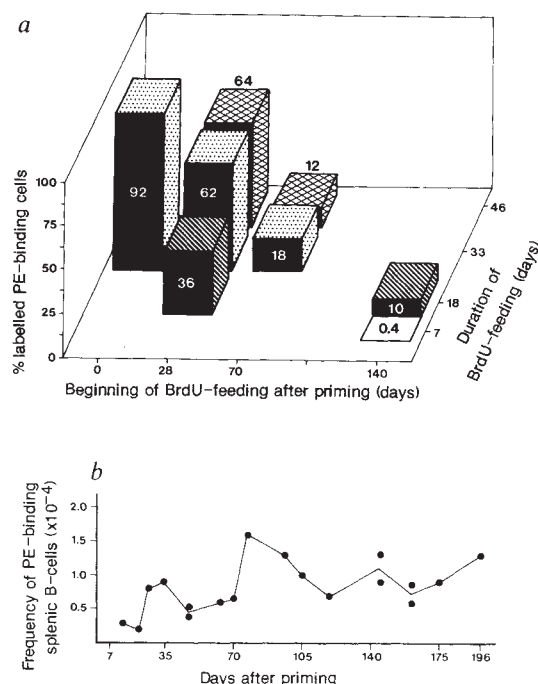


FIG. 3 Summary of the proliferative activity (a) and the frequency (b) of memory B cells in PE-immunized mice. a, Three-dimensional plot of the percentages of BrdU-labelled B220⁺PE⁺ spleen cells depending on the onset and the duration of BrdU-feeding. Each column represents a single experimental point, that is, the fraction of BrdU-labelled PE-binding cells isolated from pooled spleen cells of three mice. The position of the individual columns marks the beginning of BrdU-feeding after priming (from left to right) and the duration of BrdU-feeding (from front to back). The height of the column represents the fraction of BrdU-labelled cells among all PE-binding cells. The numbers depicted in the columns give the actual percentage of BrdU-labelled PE-binding cells. Spleen cells were analysed at the end of the labelling period as described in Fig. 1. b, Frequency of PE-binding splenic B-cells at different time points after priming. Background level of PE-binding cells in the spleen (unimmunized mice or mice primed with an unrelated antigen) is 5×10^{-6} . Each point represents the fraction of PE-binding cells in pooled spleen cells from three mice.

TABLE 1 *In vitro* secondary antibody responses of PE-binding and PE-nonbinding splenic B cells from PE-immunized mice

B cells	T cells	Anti-PE IgG1 antibodies (units)					
		Expt.	1	2a	2b	3a	3b
B220 ⁺ PE ⁺	—	*	—	—	—	—	—
B220 ⁺ PE ⁺	CD4 ⁺ cells	19	10	19	12	18	—
B220 ⁺ PE ⁻	—	—	—	—	—	—	—
B220 ⁺ PE ⁻	CD4 ⁺ cells	—	—	—	—	—	—
—	CD4 ⁺ cells	—	—	—	—	—	—

Spleen cells from PE-immunized mice were isolated 7 (experiment (expt) 2) or 12 weeks (expt 1 and 3) after priming, stained and enriched for IgM⁺IgD⁺ B cells as described in Fig. 1. B220⁺PE⁺ (expt 1, 250 cells per well; expts 2 and 3, 500 cells per well) and B220⁺PE⁻ cells (expt 1, 7×10^3 cells per well; expt 2, 4×10^4 cells per well; expt 3, 8×10^4 cells per well) were sorted into 96-well microtitre plates as described elsewhere⁷. Cells were cultured in Mishell-Dutton medium with 10% FCS either in the presence or the absence of CD4⁺ T cells of PE-immunized mice (2×10^5 cells per well). Splenic CD4⁺ T cells were obtained by positive enrichment on a MACS¹⁰ using the antibody GK1.5 (ref. 18) and were more than 90% pure. Supernatant collected 14 days after initiation of the culture was assayed for PE-binding IgG1 antibody by an enzyme linked immunosorbent assay using a biotinylated goat-anti-mouse IgG1 antiserum (Southern Biotechnology Associates) and 96-well microtitre plates coated with PE at a concentration of $5 \mu\text{g ml}^{-1}$. In expts 2 and 3, duplicate wells were analysed (a and b). The concentration of anti-PE antibodies is expressed in units ml^{-1} , 1,000 units ml^{-1} corresponding to the concentration of anti-PE antibody in a day-14 primary antiserum.

* <0.1 unit.

the persistence of B-cell memory requires the persistence of antigen, which is often stored on the surface of follicular dendritic cells for prolonged periods of time¹⁵. Such a dependence has been deduced from cell transfer experiments in which memory is rapidly lost if no antigen is added^{4,16,17}. Possible cellular competition in occupying space in the irradiated recipient makes it difficult to compare this situation with that in the intact animal. The well-known superiority of live over killed vaccines might also indicate a need of recurrent antigenic stimulation for the maintenance of memory cells. However, this effect could also result from other mechanisms, such as prolonged recruitment of memory cells into either the T- or B-cell compartment. Taking all these points together, a role for persisting antigen in maintaining the long-lived memory B-cells which have been demonstrated here remains to be established. □

Received 23 April; accepted 4 June 1990.

- Kocks, C. & Rajewsky, K. A. *Rev. Immun.* **7**, 537–559 (1989).
- Gowans, J. L. & Uhr, J. W. *J. exp. Med.* **124**, 1017–1030 (1966).
- Strober, S. *J. Immunol.* **117**, 1288–1294 (1976).
- Gray, D. & Skarvill, H. *Nature* **336**, 70–73 (1988).
- Gray, D., MacLennan, I. C. M. & Lane, P. J. L. *Eur. J. Immunol.* **16**, 641–648 (1986).
- Colle, J.-H., Truffa-Bacchi, P. & Freitas, A. A. *Eur. J. Immunol.* **18**, 1307–1314 (1988).
- Hayakawa, K., Ishii, R., Yamasaki, K., Kishimoto, T. & Hardy, R. R. *Proc. natn. Acad. Sci. U.S.A.* **84**, 1379–1383 (1987).
- Förster, I. & Rajewsky, K. *Proc. natn. Acad. Sci. U.S.A.* **87**, 4781–4784 (1990).
- Förster, I., Vieira, P. & Rajewsky, K. *Int. Immunol.* **4**, 321–331 (1989).
- Miltenyi, S., Müller, W., Weichel, W. & Radbruch, A. *Cytometry* **11**, 231–238 (1990).
- Nieuwenhuis, P. & Opstelten, D. *Am. J. Anat.* **178**, 421–435 (1984).
- MacLennan, I. C. M. & Gray, D. *Immunol. Rev.* **91**, 61–85 (1986).
- Goronzy, J., Weyand, C. M. & Fathman, C. G. *J. exp. Med.* **164**, 911–925 (1986).
- Vieira, P. & Rajewsky, K. *Int. Immunol.* (in the press). (update 7)
- Tew, J. G. & Mandel, T. E. *Immunology* **37**, 69–76 (1979).
- Celada, F. *J. exp. Med.* **124**, 1–14 (1967).
- Askonas, B. A., Cunningham, A. J., Kreth, H. W., Roelants, G. E., Williamson, A. R. *Eur. J. Immunol.* **2**, 494–498 (1972).
- Dialynas, D. P. et al. *J. Immunol.* **131**, 2445–2451 (1983).
- Schüppel, R., Wilke, J. & Weiler, E. *Eur. J. Immunol.* **17**, 739–741 (1987).
- Nishikawa, S. I., Sasaki, Y., Kina, T., Amagai, T. & Katsura, Y. *Immunogenetics* **23**, 137–139 (1986).
- Oi, V. T. & Herzenberg, L. A. *Molec. Immunol.* **16**, 1005–1017 (1979).
- Marshak-Rothstein, A. et al. *J. Immunol.* **122**, 2491–2497 (1979).
- Coffman, R. *Immunol. Rev.* **69**, 5–23 (1982).
- Khaffan, H., Abuknesha, R., Raud-Weaver, M., Price, R. G. & Robinson, D. *Histochem. J.* **18**, 497–499 (1986).
- Gonchoroff, N. J., Greipp, P. R., Kyle, R. A. & Katzmman, J. A. *Cytometry* **6**, 506–512 (1985).
- Gonchoroff, N. J. et al. *J. Immunol. Meth.* **93**, 97–101 (1986).

ACKNOWLEDGEMENTS. We are grateful to Dr N. Gonchoroff for gift of the anti-BrdU antibody BU-1, Dr W. Müller and C. Uthoff-Hachenberg for providing phycoerythrin and antibodies, C. Göttinger for expert technical help, I. Förster, W. Müller and A. Radbruch for discussion, U. Ringeisen for the graphical work, and E. Siegmund for typing the manuscript. This work was supported by the Deutsche Forschungsgemeinschaft and the Fazit Foundation.

β_2 -Microglobulin restriction of antigen presentation

Béatrice Pérarnau*, Claire-Anne Siegrist*, Anne Gillet*, Claude Vincent†, Shoji Kimura‡ & François A. Lemonnier*

* Centre d'Immunologie INSERM-CNRS de Marseille-Luminy, Case 906, 13288 Marseille Cedex 9, France

† Unité INSERM 80, 5 place d'Arsonval, 69374 Lyon Cedex 8, France

‡ Memorial Sloan-Kettering Cancer Center, 1275 York Avenue, New York 10021, USA

ANTIGENS are generally thought to be recognized by cytotoxic T lymphocytes as peptides in the context of class I major histocompatibility proteins complex^{1,2}, which are heterodimers of heavy chains noncovalently associated with β_2 -microglobulin ($\beta_2\text{m}$). The highly polymorphic nature of the heavy chains and their resulting ability to present different sets of peptides has presumably evolved to allow potent immune responses against most pathogens. By contrast, the polymorphism of $\beta_2\text{m}$ is limited; seven alleles are known in the mouse^{3,4} and only one has been identified in humans. β_2 -Microglobulin was consequently thought to have only structural functions: namely, to ensure correct folding of class I molecules and their transport to the cell surface^{5–9}. Although $\beta_2\text{m}$ is not implicated directly in the formation of the peptide binding site^{10,11}, we report here that it participates in the selection of MHC class I molecule-associated peptides.

DBA/2 (H-2^d, $\beta_2\text{m}^a$) mice were immunized with P815 (H-2^d, $\beta_2\text{m}^a$) mouse mastocytoma cells expressing the transfected mouse $\beta_2\text{m}^b$ gene. Cloned cytotoxic T lymphocytes (LAZ CTLs) showing specific lytic activity towards P815 cells transfected with mouse $\beta_2\text{m}$ were isolated. Experiments were performed to define the nature of the antigenic determinant recognized by these CTLs. They did not lyse P815 cells transfected with either the human $\beta_2\text{m}$ or a human $\times$ mouse hybrid $\beta_2\text{m}$ gene (Fig. 1), excluding, therefore, recognition of putative vector-derived T cell epitopes. Inhibition experiments with monoclonal antibodies specific for H-2K^d, D^d or L^d indicated that the eight isolated LAZ clones were H-2K^d-restricted (Fig. 2). Surprisingly, these CTLs failed to lyse H-2^{b/d}, $\beta_2\text{m}^{b/a}$ (C57BL/6 $\times$ DBA/2 F1) mouse lymphoblasts (Fig. 3a), as well as H-2^d, $\beta_2\text{m}^b$ (D2.1) B10.D2 mouse fibroblasts (Fig. 3b). These target cells were, however, efficiently lysed by control H-2K^d-restricted, HLA-Cw3-specific CAS 20 CTL clone (C.-A. Siegrist, unpublished observations).

These results led to the following hypotheses. First, the absence of lysis of C57BL/6 $\times$ DBA/2 F1 and D2.1 target cells (both expressing H-2K^d heavy chains and $\beta_2\text{m}^b$) excluded the possibility that the LAZ CTLs recognized an antigenic determinant made up of the simple association of the H-2K^d heavy chains with the mouse $\beta_2\text{m}^b$ (as suggested previously by others, in the case of H-2K^b-restricted, $\beta_2\text{m}^b$ -specific CTLs, see refs 12, 13). This conclusion was reinforced by our consistent failure to sensitize H-2^d $\beta_2\text{m}^a$ target cells to lysis by incubation with purified mouse $\beta_2\text{m}^b$ under appropriate conditions for $\beta_2\text{m}$ exchange (data not shown). Second, these results argued against the recognition of a $\beta_2\text{m}^b$ -derived peptide in an H-2K^d-restricted manner. Such a peptide should have been produced and presented by the D2.1 target cells. In some cases, however, different types of cells process internal molecules in different ways¹⁴. This argument may explain why D2.1 cells escape LAZ CTL-mediated cytotoxicity (Fig. 2b), but does not apply to P815 cells transfected with the human $\times$ mouse hybrid $\beta_2\text{m}$ gene (Fig. 1b). Considering that $\beta_2\text{m}^a$ and $\beta_2\text{m}^b$ differ only at position 85 (an alanine being replaced by an aspartic acid)¹⁵, that the usual size of the presented peptides is 10–20 amino acids and that the 30 C-terminal amino acids of the hybrid molecule are of mouse origin (Fig. 1a), one would have expected, as usually

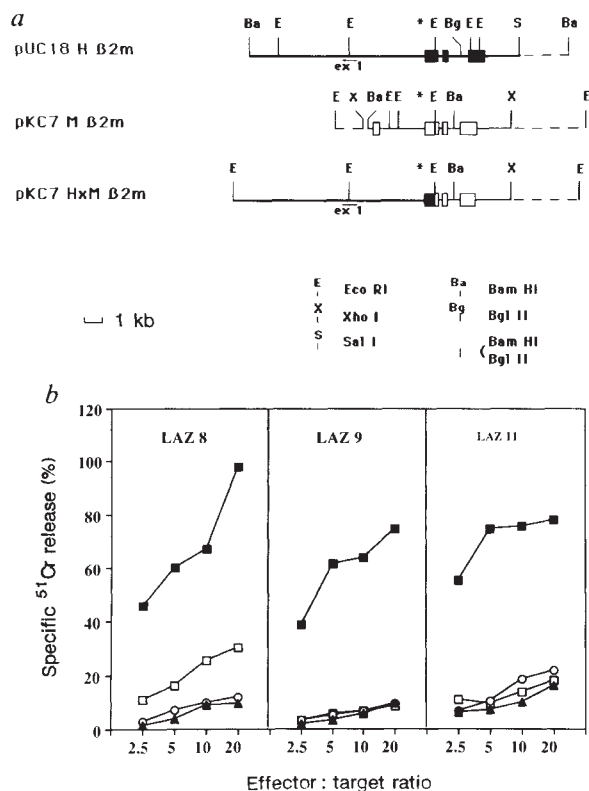


FIG. 1 Isolation of LAZ CTL clones and their reactivity on transfected P815 cells. **a**, Schematic representation of the transfected $\beta 2m$ genes. Plasmids pUC18 human $\beta 2m$ (pUC18 H $\beta 2m$; exons represented as closed boxes)²⁴ and pKC7 mouse $\beta 2m^b$ (pKC7 M $\beta 2m$, exons represented as open boxes)²⁵ were used for the construction of the hybrid $\beta 2m$ gene, taking advantage of an interspecies conserved *EcoRI* restriction site corresponding to amino acid position 70 (*E). The human $\times$ mouse hybrid $\beta 2m$ gene (pKC7 HxM $\beta 2m$) was constructed by ligation between the *EcoRI* and *XhoI* sites of the pKC7 vector of the following 3 fragments: two *EcoRI* fragments of 7 and 4.3 kilobases (kb), isolated from the pUC18 H $\beta 2m$ plasmid, encompassing the promoter, leader sequence and 5' part of exon 2 of the human $\beta 2m$ gene and an *EcoRI*-*XhoI* fragment of 4.3 kb isolated from the pKC7 M $\beta 2m$ plasmid, containing the 3' end of the mouse $\beta 2m^b$ gene. Because of the multiplicity of most restriction sites this construct required several intermediate constructions which are not represented. The approximate location of the human $\beta 2m$ first exon (ex1) is indicated. P815 HTR (H-2^d, $\beta 2m^a$, high transfection rate) mastocytoma cells were transfected with each $\beta 2m$ plasmid together with the pMCS plasmid (containing the neomycin-resistance gene). Transfected cells were selected in the presence of geneticin (GIBCO) and cloned by limiting dilution. Human $\beta 2m$ -specific BE104 ($\gamma 2b$, κ), mouse $\beta 2m^b$ -specific S19.8 ($\gamma 2b$) and IA^k-specific control 10-3.6 ($\gamma 2a$, κ) mouse monoclonal antibodies were used at saturating concentrations for indirect immunofluorescence and FACS analysis of the selected transfected clones. The means of fluorescence with the three tested monoclonals were respectively 1.6, 156, 1.5 (P815 M $\beta 2m^b$), 93, 1.7, 3.2 (P815 H $\beta 2m$) and 65, 1.9, 2.6 (P815 HxM $\beta 2m$). **b**, Isolation of LAZ CTL clones. DBA/2 (H-2^d, $\beta 2m^a$) mice were injected intraperitoneally three times with 5×10^6 P815 transfected cells expressing mouse $\beta 2m^b$ (P815 M $\beta 2m$). Spleen cells were first stimulated *in vitro* by transfected cells (30×10^6 responder cells, 1×10^6 P815 transfected cells (10,000 rad irradiated)) for 5 days and then restimulated (1×10^6 lymphoblasts from primary cultures, 10×10^6 DBA/2 feeder cells (2,000 rad irradiated)) 1×10^6 , transfected P815 cells (10,000 rad irradiated)) for 3 days. Surviving cells were subsequently cloned by seeding 4, 2 or 1 cells in round-bottomed 96-well microplates together with 10^5 feeder cells (2,000 rad irradiated) and 10^4 transfected P815 cells (10,000 rad irradiated). DMEM medium was supplemented with 10% FCS, 5×10^{-5} M 2-mercaptoethanol, 4.5 g l^{-1} glucose, 216 mg l^{-1} glutamine, 116 mg l^{-1} arginine, 36 mg l^{-1} asparagine, 10 mM HEPES buffer, 1 mM sodium pyruvate and 10% concanavalin A-stimulated rat spleen cell culture supernatant. Selected mouse $\beta 2m^b$ -specific LAZ CTLs were tested in a 4 h ^{51}Cr release assay against P815 target cells untransfected ($\square$), or transfected with human ($\blacktriangle$), mouse ($\blacksquare$), or human $\times$ mouse hybrid $\beta 2m$ ($\circ$).

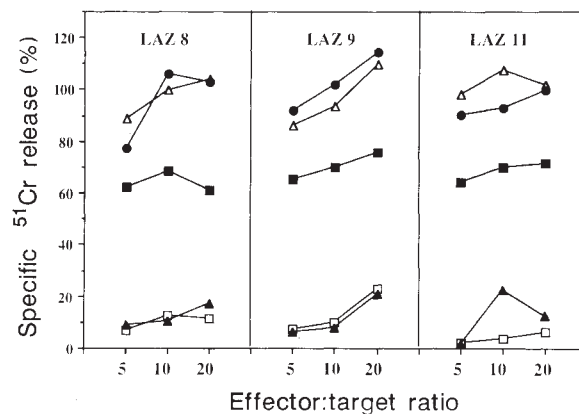


FIG. 2 H-2K^d restriction of LAZ CTLs lytic activity. Clones were tested in a 4 h ^{51}Cr release assay against untransfected P815 HTR cells ($\square$) or P815 cells transfected with mouse $\beta 2m^b$ in the absence ($\blacksquare$) or presence of 20-8-4S H-2K^d-specific ($\blacktriangle$), 34-2-12 H-2D^d-specific ($\bullet$) or 28-14-8S H-2L^d-specific ($\triangle$) monoclonal antibodies, used as unpurified ascites diluted 1:50. Results are shown for three selected (LAZ 8, LAZ 9, LAZ 11) CTL clones.

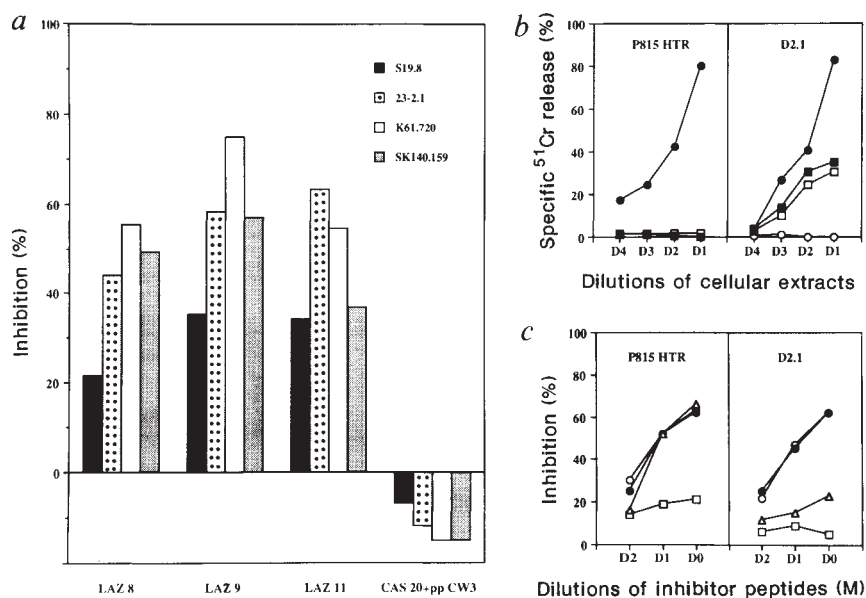
observed with fusion proteins¹⁶⁻¹⁸, that the $\beta 2m^b$ -derived peptide (which should include position 85) would have been produced and presented by the P815 cells transfected with human $\times$ mouse hybrid $\beta 2m$.

Therefore, we considered the possibility that LAZ CTLs recognize a P815 mastocytoma cell-specific peptide, whose presentation to the immune system is dependent on mouse $\beta 2m^b$. If this were the case, the peptide should be presented in exclusive association with $\beta 2m^b$ at the surface of $\beta 2m^b$ -transfected P815 (H-2^d, $\beta 2m^a$) mastocytoma cells, a P815 cell extract treated with cyanogen bromide (CNBr), containing this peptide¹⁹, should sensitize D2.1 ($\beta 2m^b$) target cells to lysis by H-2K^d-restricted LAZ CTLs and, also, one might expect some other peptides to exhibit differential affinity for the H-2K^d heavy chain depending on the associated mouse $\beta 2m$ allotype.

Experiments were performed to test these three predictions. The lytic activity of LAZ CTLs was inhibited by $\beta 2m^b$ -specific monoclonal antibodies. HLA-Cw3-specific CAS 20 CTLs were not inhibited, as expected, as the pCw3 peptide is presented to the CAS 20 clone with equal efficiency in an H-2K^d $\times$ $\beta 2m^a$ or H-2K^d $\times$ $\beta 2m^b$ context (Fig. 4a). We also found that a CNBr-treated P815 cell lysate sensitized D2.1 ($\beta 2m^b$) but not P815 ($\beta 2m^a$) cells to lysis by LAZ CTLs. Both types of cell could, however, be sensitized to lysis by H-2K^d-restricted HLA-Cw3-specific CTLs using a similar extract from P815 cells transfected with HLA-Cw3 (Fig. 4b). Finally, a series of H-2K^d-binding peptides was tested as competitive inhibitors of the fixation of the pCw3 peptide to H-2K^d. One of them, pK^d174, was a more potent inhibitor of the fixation of the pCw3 peptide to H-2K^d $\times$ $\beta 2m^a$ (P815) than H-2K^d $\times$ $\beta 2m^b$ (D2.1) molecules (Fig. 4c). Taken together, these results strongly suggest that the LAZ CTL clones recognized a naturally produced P815-specific peptide, only bound and presented by $\beta 2m^b$ -associated H-2K^d heavy chains.

The results reported here indicate that the structural polymorphism of mouse $\beta 2m$ influences (restricts) the presentation of peptides by MHC class I molecules to the immune system. Several contact residues have been identified between human $\beta 2m$ and the floor of the groove of the HLA-A2 and Aw68 class I heavy chains^{10,11,20}; they do not include position 85. To explain the restricting effect of this residue on peptide binding, one might postulate either intramolecular interactions between distantly related residues or species-associated three-dimensional polymorphism of class I molecules. Alternatively, one might consider that class I molecules are flexible and can adopt configurations other than the one defined by crystallography, which might be critical for peptide binding.

FIG. 4 $\beta 2m$ restriction of the fixation and presentation of peptides. *a*, Inhibition of LAZ CTLs by $\beta 2m^b$ -specific monoclonal antibodies. H-2K^d-restricted $\beta 2m^b$ -specific LAZ 8, LAZ 9 and LAZ 11 CTLs were tested against P815 M $\beta 2m^b$ -transfected cells in the presence of four different $\beta 2m^b$ -specific monoclonals used as unpurified ascites diluted 1:50. Similar experiments were performed using H-2K^d-restricted, HLA-Cw3-specific CAS 20 CTL, target cells being sensitized by addition of 1.8×10^{-8} M HLA-Cw3 (pCw3) derived 170–182 peptide. *b*, Sensitization of D2.1 cells to lysis by CNBr-treated cellular extracts. Extracts from 10^9 untransfected P815 (open symbols) or HLA-Cw3 $\times$ H $\beta 2m$ transfected P815 (closed symbols) cells were CNBr-treated¹⁹. Extracts were added to either P815 HTR ($\beta 2m^a$) or D2.1 ($\beta 2m^b$) H-2^d target cells at the onset of an otherwise standard ⁵¹Cr release assay, starting from a 1.25 mg ml^{-1} final protein concentration and testing serial twofold dilutions. Susceptibility to lysis was evaluated using either H-2K^d-restricted $\beta 2m^b$ -specific (squares) or H-2K^d-restricted HLA-Cw3-specific (circles) CTL clones. *c*, Peptide binding to H-2K^d $\beta 2m^a$ or $\beta 2m^b$ target cells. Lysis (in the absence of inhibitor) of either P815 HTR (51%) or D2.1 (63%) target cells at 10:1 E:T ratio by H-2K^d-restricted, HLA-Cw3-specific CTLs in the presence of pCw3 peptide (RYLKGKETLQRA 3.6×10^{-8} M final concentration (single-letter amino-acid code)), was evaluated in the presence of pB7 (RYLENGKDKLERA, $\square$), pNP



Arg⁻ (TYQRTRALVTG, $\circ$), pQ10^b (RYELGKETLLRT, $\bullet$) or pK^d 174 (RYLENGKETLLRT, Δ) inhibitor peptides, starting from a 10^{-4} M final concentration and testing serial 10-fold dilutions. Results are expressed as percentage of inhibition.

CTL CLONES	TARGET CELLS			
	B6	DBA/2	B6 $\times$ DBA/2	P815
	BLASTS	BLASTS	BLASTS	M $\beta 2m$
	H-2 ^b	H-2 ^d	H-2 ^{b/d}	H-2 ^d
LAZ 8	0	0	0	44.9
LAZ 9	0	0	0	47.8
LAZ 11	0	0	0	39.0
CAS 20-Cw3pp	0	0	0	1.6
CAS 20+Cw3pp	0	31.1	63.7	71.0

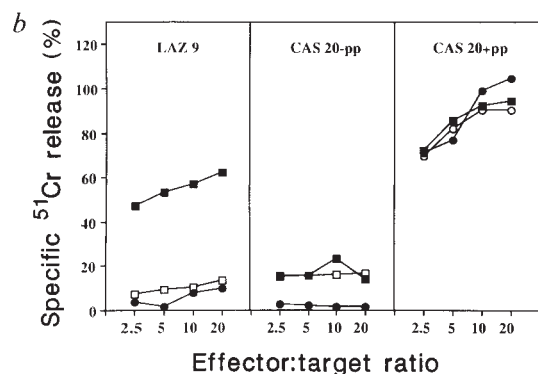


FIG. 3 Reactivity of the LAZ CTLs on different target cells. *a*, LAZ CTLs were tested at a 20:1 effector-to-target (E:T) ratio in a 4 h ⁵¹Cr release assay against C57BL/6 (B6; H-2^b, $\beta 2m^b$), DBA/2 (H-2^d, $\beta 2m^a$), C57BL/6 $\times$ DBA/2 (B6 $\times$ DBA/2; H-2^{b/d}, $\beta 2m^{a/b}$) or target cells transfected with mouse $\beta 2m^b$ (P815 M $\beta 2m$). Control H-2K^d-restricted HLA-Cw3-specific CAS 20 CTLs were tested at the same E:T ratio against the same target cells in the absence (–pp) or in the presence (+pp) of 1.8×10^{-6} M HLA-Cw3-derived peptide (amino acids 170–182). *b*, Two of the same CTLs were tested at 20:1, 10:1, 5:1 and 2.5:1 E:T ratio against untransfected ($\square$) or mouse $\beta 2m^b$ transfected ($\blacksquare$) P815 cells and H-2^d, $\beta 2m^b$ D2.1 ($\bullet$) mouse B10.D2 fibroblasts.

To what extent $\beta 2m$ restricts antigen presentation is still unknown. $\beta 2m$ restriction of antigen presentation might explain control by the H-3 region (which includes the $\beta 2m$ gene amongst others) of the induction of H-2D^b-restricted H-Y antigen-specific CTLs²¹. Presentation of some peptides, however, seems to be independent of $\beta 2m$ allotype. For example, efficient presentation of the Cw3 peptide occurs whether H-2K^d heavy chains are $\beta 2m^a$ or $\beta 2m^b$ associated (this study). Nevertheless, even a limited effect, combined with the extensive polymorphism of MHC class I heavy chains, should result in significant enlargement of the species repertoire of peptides presented by MHC class I molecules. At the same time, $\beta 2m$ polymorphism should result in the selection and presentation of different self peptides among individual mice, allowing the development of potent CTL responses and rapid allograft rejection. This could explain the strength of the H-3 minor histocompatibility locus. The reported results further indicate that the concept of minor histocompatibility antigen cannot only be restricted to the polymorphic endogenous molecules susceptible to processing into immunogenic peptides²². Other molecules²³ which may interact noncovalently with class I heavy chains might, like $\beta 2m$, indirectly influence peptide binding. $\square$

Received 9 April; accepted 6 June 1990.

1. Germain, R. N. *Nature* **329**, 512–518 (1987).
2. Townsend, A. R. M. *et al. Cell* **44**, 959–968 (1986).
3. Robinson, P. J., Steinmetz, M., Moriawaka, K. & Fischer-Lindahl, K. *Immunogenetics* **20**, 655–665 (1984).
4. Gasser, D. L., Klein, K. A., Choi, E. & Seidman, J. G. *Immunogenetics* **22**, 413–416 (1985).
5. Ploegh, H. L., Cannon, E. & Strominger, J. L. *Proc. natn. Acad. Sci. U.S.A.* **76**, 2273–2277 (1979).
6. Sege, K. L., Rask, L. & Peterson, P. A. *Biochemistry* **20**, 4523–4530 (1981).
7. Arce-Gomez, B., Jones, E. A., Barnstable, C. J., Solomon, E. & Bodmer, W. F. *Tissue Antigens* **11**, 96–112 (1983).
8. Fellous, M. *et al. Immunogenetics* **5**, 423–436 (1977).
9. Seong, R. H., Clayberger, C. A., Krensky, R. M. & Parnes, J. R. *J. exp. Med.* **167**, 288–299 (1988).
10. Bjorkman, P. J. *et al. Nature* **329**, 506–512 (1987).
11. Bjorkman, P. J. *et al. Nature* **329**, 512–518 (1987).
12. Rammensee, H.-G., Robinson, P. J., Crisanti, A. & Bevan, M. J. *Nature* **319**, 268–271 (1986).
13. Kurtz, M. E., Martin-Morgan, D. & Graff, R. J. *J. Immun.* **138**, 87–90 (1987).
14. Marrack, P. & Kappler, J. *Nature* **332**, 840–843 (1988).
15. Ramanathan, L., Dubois, G. C., Robinson, E. A. & Appella, E. *Molec. Immun.* **19**, 435–446 (1982).
16. Yamada, A., Young, J. F. & Ennis, F. A. *J. exp. Med.* **162**, 1720–1725 (1985).
17. Townsend, A. R. M., Bastin, J., Gould, K. & Brownlee, G. G. *Nature* **324**, 575–577 (1986).
18. Chimini, G., Pala, P., Sire, J., Jordan, B. R. & Maryanski, J. L. *J. exp. Med.* **169**, 297–302 (1989).
19. Heath, W. R., Hurd, W. R., Carbone, F. R. & Sherman, L. A. *Nature* **341**, 749–752 (1989).
20. Garrett, P. J., Saper, M. A., Bjorkman, P. J., Strominger, J. L. & Wiley, D. C. *Nature* **342**, 692–696 (1989).

21. Fierz, W., Farmer, G. A., Sheena, J. H. & Simpson, E. *Immunogenetics* **16**, 593-601 (1982).
 22. Wallny, H.-J. & Rammensee, H.-G. *Nature* **343**, 275-278 (1990).
 23. Fehlmann, M. *et al. Proc. natn. Acad. Sci. U.S.A.* **68**, 8634-8638 (1985).
 24. Pérarnau, B. M., Gillet, A., Hakem, R., Barad, M. & Lemonnier, F. A. *J. Immun.* **141**, 1383-1389 (1988).
 25. Margulies, D. H., Parnes, J. R., Johnson, N. A. & Seidman, J. G. *Proc. natn. Acad. Sci. U.S.A.* **80**, 2328-2331 (1983).

ACKNOWLEDGEMENTS. We thank Drs T. Rajan, R. Zinkernagel and E. Larsson-Scird for providing the mouse $\beta 2m^p$ gene, D2.1 cells and pNP Arg peptide respectively, P. Ferrier, P. Le Bouteiller and L. Leserman for careful reading of the manuscript, and P. Naquet and M. Pla for useful discussions. This work was supported by the Institut National de la Recherche Médicale, the Centre National de la Recherche Scientifique, the Association pour la Recherche contre le Cancer and Fonds National de la Recherche Scientifique Suisse.

Suppression of *c-ras* transformation by GTPase-activating protein

Ke Zhang*, Jeffrey E. DeClue*, William C. Vass*, Alex G. Papageorge*, Frank McCormick† & Douglas R. Lowy*

* Laboratory of Cellular Oncology, National Cancer Institute, Bethesda, Maryland 20892, USA

† Department of Molecular Biology, Cetus Corporation, Emeryville, California 94608, USA

THE *ras* genes are required for normal cell growth and mediate transformation by oncogenes encoding protein tyrosine kinases^{1,2}. Normal *ras* can transform cells *in vitro* and *in vivo*, but mutationally activated *ras* does so much more efficiently, and highly transforming mutant versions of *ras* have been isolated from a variety of human and animal tumours (reviewed in refs 3-5). The *ras* genes encode membrane-associated, guanine nucleotide-binding proteins that are active when GTP is bound and inactive when GDP is bound⁶⁻⁸. The slow intrinsic GTPase activity of normal mammalian Ras proteins can be greatly accelerated by the GTPase-activating protein (GAP), which is predominantly cytoplasmic. This activity of GAP, which can increase with cell density in contact-inhibited cells⁹, suggests that it functions as a negative, upstream regulator of *ras*. Other studies^{10,11}, however, show that GAP interacts with a region of *ras*-encoded protein implicated in *ras* effector function^{12,13}, which raises the possibility that GAP might also be a downstream target of *ras* (reviewed in ref. 14). Mutationally activated *ras*-encoded proteins also interact with GAP¹⁵, although they are resistant to its catalytic activity⁷. In an attempt to define the role of GAP in *ras*-mediated transformation, we examined the effects on transformation of normal or mutant *ras* when cells overexpress GAP. We found that GAP suppresses transformation of NIH 3T3 cells by normal Ha-*ras* (*c-ras*) but does not inhibit transformation by activated Ha-*ras* (*v-ras*). These results support the hypothesis that GAP functions as a negative regulator of normal *ras* and make it unlikely that GAP alone is the *ras* target.

NIH 3T3 cells were transfected with normal or mutant *ras* genes placed under the control of a murine retroviral long terminal repeat. The human GAP complementary DNA clone¹⁶, encoding the full-length 1,047-amino-acid GAP protein, was expressed from a Moloney murine leukaemia virus long terminal repeat in the mammalian expression vector pGV16 (refs 17, 18) (Fig. 1). Transfection of the NIH 3T3 line with normal *c-ras* resulted in the formation of a few dozen to over a hundred foci (Table 1). The presence of a three- to six-fold excess of pGV16 vector either bearing no inserted DNA (pKZ93) or with the GAP cDNA in the antisense orientation (pKZ75) did not appreciably inhibit the number of foci obtained with *c-ras* (Table 1; data not shown). However, when the plasmid pKZ61, encoding GAP in the sense orientation, was cotransfected with *c-ras* in a three- to six-fold excess, a striking inhibition in the number of foci was observed. The average inhibition in four independent

TABLE 1 Inhibition of *c-Ras*-induced foci by cotransfection with cloned GAP

	Experiment				
	1	2	3	4	Ave.
<i>c-Ras</i> /GAP	1	7	40	11	
<i>c-Ras</i> /control 1	28	—	96	89	
<i>c-Ras</i> /control 2	—	134	—	—	
% inhibition by GAP	96	95	58	88	84
<i>v-Ras</i> /GAP	530	268	208	151	
<i>v-Ras</i> /control 1	508	—	242	161	
<i>v-Ras</i> /control 2	—	394	—	—	
% inhibition by GAP	-4	32	16	6	13
<i>v-Mos</i> /GAP	314				
<i>v-Mos</i> /control 1	214				
% inhibition by GAP	-47				-47

GAP, pKZ61; control 1, pKZ93; control 2, pKZ75 (see Fig. 1). Transfections were carried out and foci scored as described previously⁵. In the various experiments, 25-100 ng per dish of Ras was used with 100-300 ng GAP plasmid or control plasmid. In experiments 1, 2 and 4, the Ras to GAP ratio was 1:3. In experiment 3, the ratio was 1:6. Ave., average.

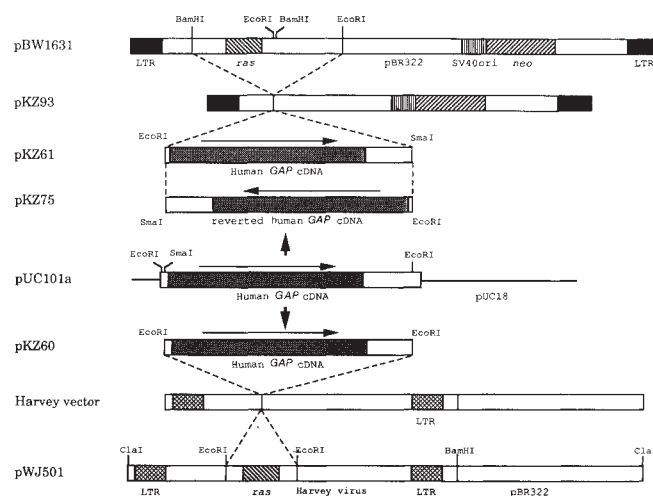
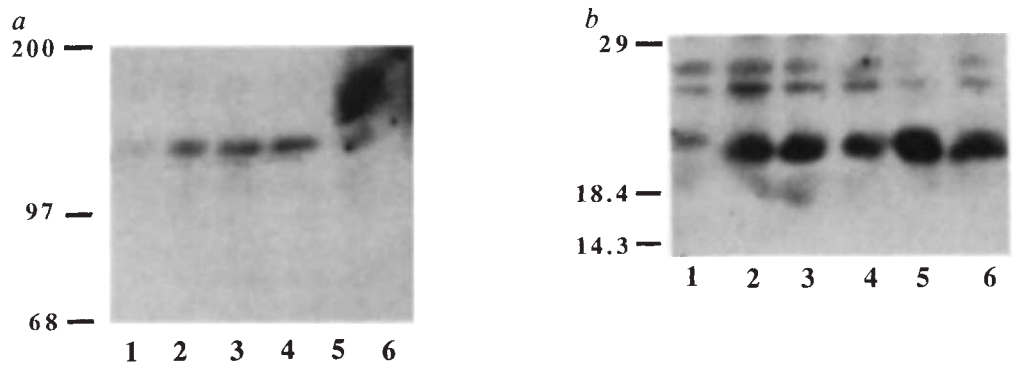


FIG. 1 Structure of plasmids with GAP and *ras*.

METHODS. Plasmid pBW1631, which contains *c-ras*, has been described previously¹⁸. The activated *v-ras* oncogene of Harvey murine sarcoma virus differs from *c-ras* only at codon 12 (encoding arginine instead of glycine) and codon 59 (encoding threonine instead of glutamine). Plasmid pKZ93, the vector without an insert, was constructed by digesting pBW1631 with *Bam*HI and *Eco*RI; the ends were made flush with Klenow fragment and religated. For pKZ61 and pKZ75, a full-length human GAP cDNA, from clone pUC101a, was digested with *Eco*RI and *Sma*I, the ends were made flush, and then ligated to the flush ended pKZ93 precursor. Plasmid pKZ61 contains the GAP cDNA in the sense orientation, pKZ75 in the antisense orientation. Plasmid pKZ60, which is a second GAP cDNA clone in the sense orientation, was constructed by substituting a GAP *Eco*RI fragment from pUC101a for the *v-ras* *Eco*RI fragment in pWJ1501. The construction of pWJ1501 will be described elsewhere (W. M. Ju *et al.*, manuscript in preparation). It is Harvey murine sarcoma virus DNA cloned in pBR322 that has been modified to contain only two *Eco*RI sites, just upstream and downstream of *v-ras*, by substituting an *Eco*RI site for *Sac*II at nucleotide 940, substituting an *Eco*RI polylinker at the *Pst*I site at nucleotide 1,759, and removing the *Eco*RI site at nucleotide 2,540.

transfections was 84% (Table 1). In contrast, the number of foci produced by the activated *v-ras* gene, which consistently results in a higher number of foci than *c-ras*, was not significantly affected by cotransfection with pKZ61. There was no inhibition of focus formation by *v-mos*¹⁹, whose transforming activity is *ras*-independent², when it was cotransfected with the GAP-encoding plasmid (Table 1).

FIG. 2 Expression of GAP and Ras proteins in revertant and nonrevertant subclones. *a*, Level of expression of GAP protein as detected by western blotting with anti-GAP serum. Lane 1, normal NIH 3T3; lanes 2–4, revertant subclones 8–1, 8–2 and 8–4, respectively; lane 5, non-revertant subclone 5-1T (isolated from *c-ras*-transformed cells transfected with the selectable marker alone); lane 6, nonrevertant subclone 8–10T (isolated from the same population as the revertant subclones). Arrow indicates GAP protein. *b*, Expression of Ha-*c-ras* p21 in cell lines as detected by western blotting with the anti-Ras monoclonal antibody 259. Lane 1, normal NIH 3T3; lanes 2–4, revertant subclones 8–1, 8–2 and 8–4, respectively; lane 5, nonrevertant subclone 5–1T; lane 6, nonrevertant subclone 8–10T. Arrow indicates Ras p21 protein.



METHODS. *a*, Confluent 100-mm plates were lysed in 28 mM Tris buffer, pH 7.1, 100 mM NaCl, 5 mM MgCl₂, 1% NP-40, 0.5% sodium deoxycholate, 1 mM DTT and 16 µg ml⁻¹ aprotinin. Protein content of the lysates was determined using the bicinchoninic acid assay³¹ (Pierce Chemical Co.), and 100 µg of each lysate was electrophoresed in an 8% SDS-polyacrylamide

gel. Proteins were electroblotted onto an Immobilon membrane (Millipore Corp.), incubated with a polyclonal rabbit anti-GAP serum, then a goat anti-rabbit serum (Organon Teknika Corp.), and 10⁷ c.p.m. ¹²⁵I-conjugated protein A (NEN Dupont Co.). Autoradiography was for 7 days at -70 °C. Migration of molecular weight standards is shown on the left (*M_r* × 10⁻³). *b*, Western blotting was carried out as described for *a*, except that 20 µg protein was separated on a 15% polyacrylamide gel and the antibody incubations were with the rat monoclonal Y13–259 (ref. 32) followed by a rabbit anti-rat serum, then ¹²⁵I-conjugated protein A. Autoradiography was for 3 days at -70 °C. Migration of *M_r* standards is shown on the left (*M_r* × 10⁻³).

These data demonstrate that *GAP* can prevent transformation by *c-ras* but not by the activated *v-ras* allele. We did not see potentiation of *c-ras* transformation by *GAP*, as might have been expected if *GAP* alone were the *ras* target. The results raised the possibility that overexpression of *GAP* protein might be capable of inducing morphological reversion of cells already transformed by *c-ras*, but not by *v-ras*.

To test this idea, two NIH 3T3 lines transformed by *c-ras* and one *v-ras*-transformed line were cotransfected with an MuLV-based vector carrying the human *GAP* cDNA (Fig. 1, pKZ60) and a selectable marker, pSV2his or pSV2neo^r, conferring resistance to L-histidinol²⁰ or geneticin, respectively. Transfection of both *c-ras* lines yielded a large proportion of morphologically revertant resistant colonies when transfected with the *GAP* cDNA, and only a few when cotransfected with the control plasmid (Table 2). In contrast, there was no reversion of the *v-ras*-transformed line by the *GAP* cDNA.

TABLE 2 Morphological reversion of transformation by *GAP*

	Transformed cell lines	
	% of flat colonies (no. revertants/total)	
	<i>c-ras</i>	<i>v-ras</i>
Experiment 1		
Control	5 (3/56)	ND
GAP	20 (15/74)	ND
Experiment 2		
Control	1 (1/84)	1 (1/146)
GAP	85 (17/20)	2 (1/48)

Control, insertless Harvey vector; GAP, pKZ60 (see Fig. 1). One microgram of GAP or control DNA was cotransfected with 100 ng pSV2his plasmid in experiment 1. The same DNAs were cotransfected with 100 ng pSV2 neo in experiment 2. The *c-ras*-transformed clone used in experiment 1 was obtained by transfection with pBW1631, which contains *neo^r* as a linked marker, selection of geneticin-resistant colonies, and cloning of a transformant by end-point dilution. The *c-ras*-transformed clone used in experiment 2 was obtained by transfection of pC07-FX (ref. 29), in which *c-ras* is under control of the Friend MuLV LTR, and selection of a transformant by end-point dilution. The *v-ras*-transformed clone was obtained by transfection with pKZ132 (in which the *v-ras* coding sequences have been inserted in the unique *Bgl*II site of pKSV-10 (ref. 30) by blunt-end ligation) and cloning by end-point dilution.

To confirm that the reversion of *c-ras* transformation was indeed due to expression of exogenous *GAP*, several drug-resistant subclones were isolated from among those analysed (Table 2, experiment 1). Northern blot analysis revealed that the revertant, but not the transformed, subclones expressed high levels of exogenous *GAP* RNA (data not shown). Cell extracts from the subclones were analysed by western blotting for *GAP* protein (Fig. 2*a*), which is predominantly cytoplasmic⁷, although a proportion of *GAP* that has been phosphorylated on tyrosine in response to stimulation by platelet-derived growth factor is membrane-associated²¹. Under these assay conditions, the endogenous *GAP* protein of relative molecular mass 120,000 was barely detectable in NIH 3T3 cells (lane 1) or in *c-ras* transformants (lanes 5 and 6), whereas the revertants, as expected, contained much higher levels of *GAP* protein (lanes 2–4). These results confirm that the cells that were morphologically reverted by transfection of the *GAP* cDNA expressed significant levels of exogenous *GAP* protein. To exclude the possibility that the *GAP*-induced reversion might have resulted from a loss of *c-ras* expression, the cell lysates from the same set of subclones were examined by western blotting for the p21 protein encoded by *ras* with a Ras-specific monoclonal antibody. Parental NIH 3T3 cells expressed only a low level of p21, as expected (Fig. 2*b*, lane 1). In contrast, the five subclones derived from the *c-ras*-transformed line expressed abundant levels of p21 (lanes 2–6), whether they were morphologically transformed or revertant. Thus the level of *GAP* expression, rather than p21, correlates with the nontransformed state of the revertant subclones.

The findings presented here show that *GAP* can inhibit and reverse *c-ras* transformation of mammalian cells, whereas *GAP* had no detectable effect on transformation by the activated *v-ras* gene. The reversion of *c-ras* transformation by *GAP* indicates that an essential part of the *ras* transformation pathway in mammalian cells is overridden by the introduction of excess *GAP* into the cells. Thus the present results lend strong support for the role of *GAP* as an upstream negative regulator of *ras*, with *GAP* presumably functioning, through its catalytic activity, to reduce the level of active GTP-bound Ras protein. Consistent with this interpretation, a *GAP*-derived clone encoding only the C-terminal 340 amino acids, which constitutes the enzymatic domain of *GAP* protein²², was also capable of suppressing *c-ras* transformation in these assay systems (unpublished observation). The results suggest that *GAP* as a negative regulator is

limiting in NIH 3T3 cells and raise the possibility that overexpression of GAP might inhibit transformation by *ras*-dependent oncogenes, such as membrane associated tyrosine kinases².

If GAP alone represented the *ras* target, overexpression of GAP would be expected either to potentiate transformation or to have no influence on it. The inhibition of transformation by the presence of excess GAP is in stark contrast to the demonstration that overexpression of several proto-oncogenes, including *c-ras*, can induce transformation of established cells. The results indicate that GAP can act independently as a negative regulator of Ras activity, regardless of its possible downstream function.

The results obtained in mammalian cells are analogous to those recently reported for the biological activity of mammalian GAP in yeast²³. In that study, mammalian GAP inhibited the function of *c-ras* but not mutationally activated *ras*. In addition, GAP functionally complemented yeast strains defective for the function of the *IRA1* and *IRA2* genes^{23,24}, which are upstream negative regulators of yeast RAS. But the study of GAP in yeast may not have allowed for an unbiased assessment of the role of GAP and its interaction with *ras*. GAP is not in question as the RAS target in yeast, and the known effector of RAS in yeast is adenylyl cyclase²⁵, which does not seem to be a target for Ras in higher eukaryotes^{26,27}.

Although our results tend to rule out the simple situation of GAP alone being the Ras target, they may not exclude the possibility that GAP might, in addition to being a negative regulator of Ras, participate in Ras target function if the target were a complex that involved molecules in addition to GAP. In this case, overexpression of GAP might inhibit *ras* transformation if a delicate stoichiometric balance required for the proper functioning of the hypothetical complex were disrupted by the presence of excess GAP in the cells. It may be relevant that GAP has recently been found to form a complex with two proteins that are not Ras in cells transformed by a variety of oncogenes that encode tyrosine protein kinases²⁸. But the potential physiological significance of this association might be limited to GAP's function as an attenuator, and such complexes have not so far been identified in *ras*-transformed cells. Further analysis is required for a more complete understanding of the role of the Ras-GAP interaction in signal transduction. □

Received 13 March; accepted 16 July 1990.

- Mulcahy, L. S., Smith, M. R. & Stacey, D. W. *Nature* **313**, 241–243 (1985).
- Smith, M. R., DeGudicibus, S. J. & Stacey, D. W. *Nature* **320**, 540–543 (1986).
- Barbacid, M. *Ann. Rev. Biochem.* **56**, 779–827 (1987).
- Gibbs, J. B. & Marshall, M. S. *Microbiol. Rev.* **53**, 171–185 (1989).
- Velu, T. J., Vass, V. C., Lowy, D. R. & Tambourin, P. E. *J. Virol.* **63**, 1384–1392 (1989).
- Field, J., Broek, D., Kataoka, T. & Wigler, M. *Molec. cell. Biol.* **7**, 2128–2133 (1987).
- Trahey, M. & McCormick, F. *Science* **238**, 542–545 (1987).
- Satoh, T., Nakamura, S. & Kaziro, Y. *Molec. cell. Biol.* **7**, 4553–4556 (1987).
- Hoshino, M., Kawakita, M. & Hattori, S. *Molec. cell. Biol.* **8**, 4169–4173 (1988).
- Adari, H., Lowy, D. R., Willumsen, B. M., Der, C. J. & McCormick, F. *Science* **240**, 518–521 (1988).
- Calés, C., Hancock, J. F., Marshall, C. & Hall, A. *Nature* **332**, 548–551 (1988).
- Sigal, I. S., Gibbs, J. B., D'Alonzo, J. S. & Scolnick, E. M. *Proc. natn. Acad. Sci. U.S.A.* **83**, 4725–4729 (1986).
- Willumsen, B. M. *et al. Molec. cell. Biol.* **9**, 2646–2654 (1986).
- McCormick, F. *Cell* **56**, 5–8 (1989).
- Vogel, U. S. *et al. Nature* **335**, 90–93 (1988).
- Trahey, M. *et al. Science* **242**, 1697–1700 (1988).
- Jhappan, C., Vande Woude, G. P. & Robins, T. J. *Virology* **60**, 750–753 (1986).
- Stone, J. C., Vass, W. C., Willumsen, B. M. & Lowy, D. R. *Molec. cell. Biol.* **8**, 3565–3569 (1988).
- Blair, D. G., McClements, W. L., Oskersson, M. K., Fischinger, P. J. & Vande Woude, G. F. *Proc. natn. Acad. Sci. U.S.A.* **77**, 3504–3508 (1980).
- Hartman, S. C. & Mulligan, R. C. *Proc. natn. Acad. Sci. U.S.A.* **85**, 8047–8051 (1988).
- Molloy, C. J. *et al. Nature* **342**, 711–714 (1989).
- Marshall, M. S. *et al. EMBO J.* **8**, 1105–1110 (1989).
- Ballester, R. *et al. Cell* **59**, 681–686 (1989).
- Tanaka, K. *et al. Cell* **60**, 803–807 (1990).
- Field, J. *et al. Molec. cell. Biol.* **8**, 2159–2165 (1988).
- Beckner, S. K., Hattori, S. & Shih, T. Y. *Nature* **317**, 71–73 (1985).
- Levitzi, A., Rudick, J., Pastan, I., Vass, W. C. & Lowy, D. R. *FEBS Lett.* **197**, 134–138 (1986).
- Ellis, C., Moran, M., McCormick, F. & Pawson, T. *Nature* **343**, 377–381 (1990).
- Feldman, R. A., Lowy, D. R., Vass, W. C. & Velu, T. J. *J. Virol.* **63**, 5469–5474 (1989).
- Dhruva, B. R., Shenk, T. & Subramanian, K. N. *Proc. Acad. Sci. U.S.A.* **77**, 4514–4518 (1980).
- Smith, P. K. *et al. Anal. Biochem.* **150**, 76–85 (1985).
- Furth, M. E., Davis, L. J., Fleurdelys, B. & Scolnick, E. M. *J. Virol.* **43**, 294–304 (1982).

ACKNOWLEDGEMENTS. We thank Berthe Willumsen, Patrick Martin and John Schiller for helpful discussions.

Obligatory wounding requirement for tumorigenesis in *v-jun* transgenic mice

Andre C. Schuh*, Sarah J. Keating†‡, Felipe S. Monteclaro§, Peter K. Vogt§ & Martin L. Breitman*||

* Division of Molecular and Developmental Biology, Samuel Lunenfeld Research Institute and † Department of Pathology, Mount Sinai Hospital, 600 University Avenue, Toronto, Ontario M5G 1X5, Canada
Departments of ‡ Pathology and || Medical Genetics, University of Toronto, Toronto, Ontario M5S 1A8, Canada
§ Department of Microbiology, USC School of Medicine, Los Angeles, California 90033-1054, USA

AVIAN sarcoma virus 17 induces fibrosarcomas in chickens¹ and can transform a number of avian cell types *in vitro*² by the action of *v-jun* (refs 1–3). This gene and the related cellular genes *c-jun*^{4–6}, *jun B*⁷ and *jun D*^{8,9}, encode transactivating (or repressing) DNA-binding proteins that form homo- or heterodimeric (Jun–Jun and Jun–Fos) complexes^{3,10,11} which recognize the AP-1 consensus sequence TGACTCA (refs 12, 13), a response element that confers sensitivity to the tumour-promoting phorbol ester TPA¹⁴. We have produced several lines of transgenic mice carrying the *v-jun* oncogene, driven by the promoter of the widely expressed H-2K^k major histocompatibility complex (MHC) class I antigen gene. Transgenic animals are initially phenotypically normal, but after full-thickness wounding they show abnormal wound repair, characterized by hyperplastic granulation tissue. Many of these lesions are slowly progressive because of continuing fibroblast proliferation, and over 2–5 months some give rise to dermal fibrosarcomas. This reproducible multistep transition through a proliferative but benign intermediate is associated with characteristic increments in *v-jun* expression. Moreover, hyperplastic wound repair and its progression are both related to transgene dosage, suggesting that there exists a quantitative requirement or threshold for *v-jun* action. Our results indicate that *v-jun* is not oncogenic in transgenic mice as a result of a 'single-hit' mechanism, but rather, in addition to an obligatory wound, that secondary genetic or epigenetic events (possibly conscripting normal constituents of wound repair) are necessary for tumour development and progression.

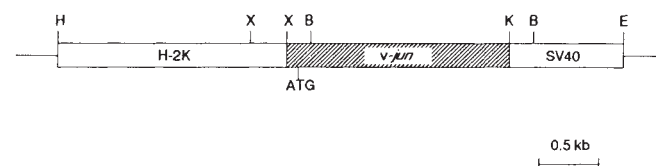


FIG. 1 Structure of the H-2K-*v-jun* transgene. The *gag-jun* fusion gene from ASV17 is driven by a 1.9-kb fragment from the 5' end of the murine H-2K^k gene, ending 16 base pairs (bp) upstream of the H-2K translation initiation site. The polyadenylation signal is derived from the early region of SV40. Translation of *v-jun* mRNA begins at the initiation codon of the *gag* gene. H, HindIII; X, XbaI; B, BclI; K, KpnI; E, EcoRI sites, respectively. ATG, *gag* initiation site.

METHODS. A 988-bp *EcoRI*–*BclI* fragment from SV40 was inserted between the *EcoRI* and *BclI* sites of a plasmid pSP64 derivative in which the *XbaI* site had been converted to a *KpnI* site. The SV40 fragment was then retrieved as an *EcoRI*–*KpnI* fragment, and inserted between the *EcoRI* and *KpnI* sites of pG5-5-1, a pGEM4 derivative containing a 1.8 kb *v-jun* fragment from ASV 17 (ref. 1) cloned into the *SmaI* site, producing *pv-junPA*. The 1.9-kb H-2K^k promoter segment¹⁵ was isolated from pH2K^k as a *HindIII*–*XbaI* fragment and inserted between the *HindIII* and *XbaI* sites of the *pv-junPA* intermediate to give pH-2K-*v-jun*. The final construct was digested with *HindIII* and *EcoRI* and a 4.7-kb fragment purified for microinjection into fertilized mouse eggs as described³³.

TABLE 1 Relative *v-jun* expression in H-2K-*v-jun* transgenic lines

Line	Th	Sp	H	Lu	Li	K	Br	Sk	Mu	Bo	Te	O	U	EF
Tg1	++++	+++	+	+++	+/-	+++	+/-	++	-	-	++	++++	++++	ND
Tg2	++++	+++	-	+/-	+/-	+/-	+/-	++	+/-	+/-	+++	-	-	++
Tg4	+++	+++	-	++	-	++	+	++	+/-	+/-	45+	ND	ND	ND

Steady-state levels of *v-jun* mRNA in a variety of transgenic tissues from hemizygous animals are shown. Tg1, Tg2 (C3H/C57Bl/6 hybrid background) and Tg4 (CD-1 background) contain respectively ~10, 150 and 50 copies of the transgene. The symbols +, ++, +++, +++++, represent linear increments in *v-jun* mRNA levels; +/-, trace; -, negative; ND, not done; Th, thymus; Sp, spleen; H, heart; Lu, lung; Li, liver; K, kidney; Br, brain; Sk, skin; Mu, skeletal muscle; Bo, bone; Te, testis; O, ovary; U, uterus; EF, homozygous primary embryo fibroblasts. In parallel experiments, *v-jun* expression was not detected in any non-transgenic tissues.

For northern analysis, 15 µg total RNA, prepared by the guanidinium isothiocyanate-caesium chloride method, was separated on a 1% agarose-formaldehyde gel, transferred to nitrocellulose and hybridized at 42 °C as previously described³² to a random-primed *v-jun* probe (encompassing the entire 1.8-kilobase *gag-jun* fusion gene) in 25 mM KPO₄, pH 7.4, 5 × SSC, 5 × Denhardt's solution, 50% formamide, 50 µg ml⁻¹ salmon sperm DNA and 10% dextran sulphate. Blots were first washed in 2 × SSC, 0.1% SDS at 55 °C, and then sequentially in 0.2 × SSC, 0.1% SDS at 55 °C and 65 °C. Quantitation of signal intensity in equally loaded lanes was by laser densitometry using a Molecular Dynamics 300A computing densitometer.

To investigate the oncogenicity of *v-jun* in multiple cellular lineages, we have generated several lines of transgenic mice carrying *v-jun* driven by the promoter of the widely expressed H-2K^b MHC class I antigen gene¹⁵ (Fig. 1). Two lines of these H-2K/*v-jun* transgenic mice (Tg1 and Tg2) and an additional sterile founder (Tg4) have been studied in detail (Table 1). Constitutively expressed *v-jun* messenger RNA is detectable in varying amounts in a wide range of adult tissues, but with some differences in corresponding tissues among independent transgenic lines (Table 1; Fig. 2a). In general, *v-jun* expression is not related to transgene copy number, and there is a three- to fourfold difference in *v-jun* mRNA levels between high-expressing tissues such as thymus, spleen and testis and low-expressing tissues such as heart, liver, brain, muscle and bone.

H-2K-*v-jun* animals show no abnormalities in gross phenotype and a wide range of tissues from Tg1, Tg2 and Tg4 have been microscopically normal as well. A notable exception was the testes of the sterile Tg4, which were much smaller than those of control animals and microscopically showed a block to spermatogenesis at roughly the pachytene primary spermatocyte stage (not shown). The levels of *v-jun* mRNA in the testes of Tg4 were at least 15-fold higher than in those of fertile Tg1 and Tg2 mice (Table 1), suggesting that *v-jun* at comparatively high levels may block spermatogenesis. After a mean observation time of 8.5 months (up to 17 months in some cases), animals carrying H-2K-*v-jun* do not differ in spontaneous tumour incidence from non-transgenic strain-matched animals.

Although the transgenic animals are initially phenotypically

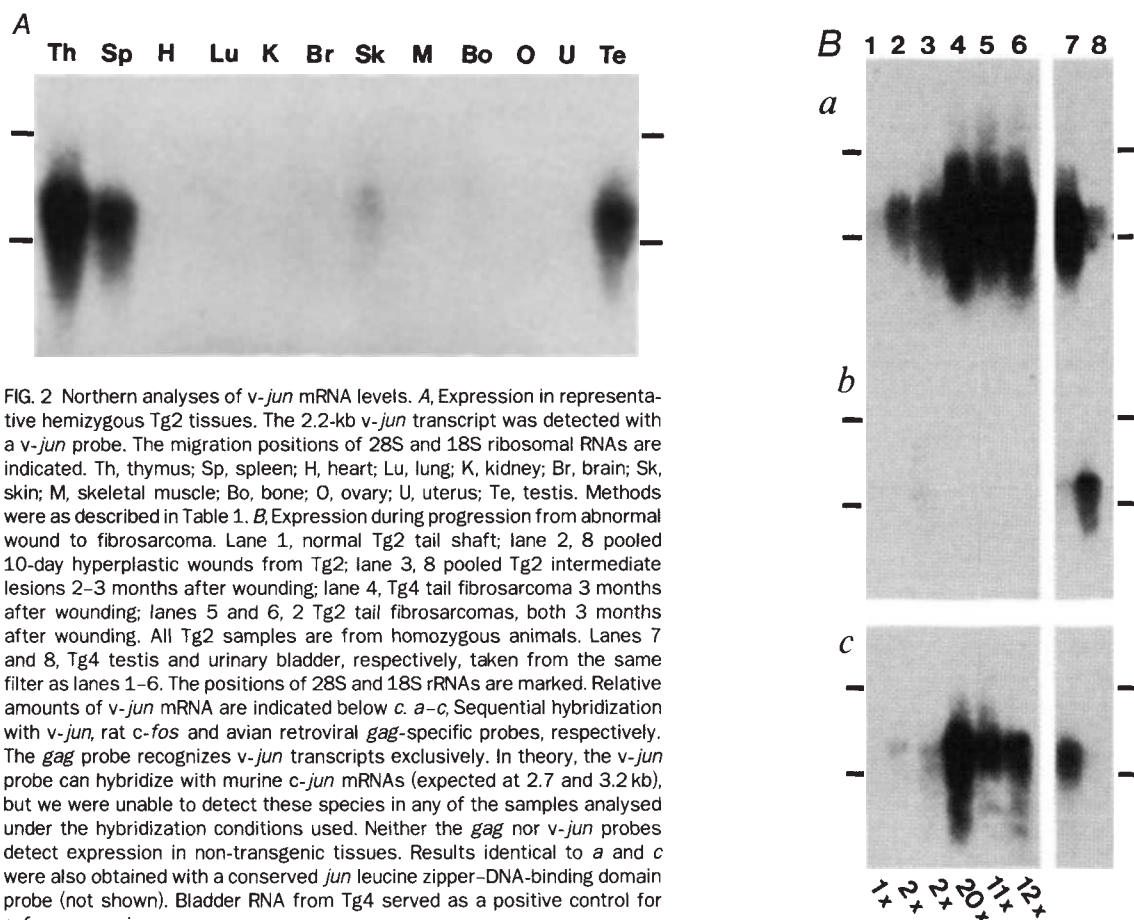


FIG. 2 Northern analyses of *v-jun* mRNA levels. **A**, Expression in representative hemizygous Tg2 tissues. The 2.2-kb *v-jun* transcript was detected with a *v-jun* probe. The migration positions of 28S and 18S ribosomal RNAs are indicated. Th, thymus; Sp, spleen; H, heart; Lu, lung; K, kidney; Br, brain; Sk, skin; M, skeletal muscle; Bo, bone; O, ovary; U, uterus; Te, testis. Methods were as described in Table 1. **B**, Expression during progression from abnormal wound to fibrosarcoma. Lane 1, normal Tg2 tail shaft; lane 2, 8 pooled 10-day hyperplastic wounds from Tg2; lane 3, 8 pooled Tg2 intermediate lesions 2–3 months after wounding; lane 4, Tg4 tail fibrosarcoma 3 months after wounding; lanes 5 and 6, 2 Tg2 tail fibrosarcomas, both 3 months after wounding. All Tg2 samples are from homozygous animals. Lanes 7 and 8, Tg4 testis and urinary bladder, respectively, taken from the same filter as lanes 1–6. The positions of 28S and 18S rRNAs are marked. Relative amounts of *v-jun* mRNA are indicated below **c**. **a–c**, Sequential hybridization with *v-jun*, rat *c-fos* and avian retroviral *gag*-specific probes, respectively. The *gag* probe recognizes *v-jun* transcripts exclusively. In theory, the *v-jun* probe can hybridize with murine *c-jun* mRNAs (expected at 2.7 and 3.2 kb), but we were unable to detect these species in any of the samples analysed under the hybridization conditions used. Neither the *gag* nor *v-jun* probes detect expression in non-transgenic tissues. Results identical to **a** and **c** were also obtained with a conserved *jun* leucine zipper-DNA-binding domain probe (not shown). Bladder RNA from Tg4 served as a positive control for *c-fos* expression.

METHODS. RNA was prepared and analysed as in Table 1. **a**, The *v-jun* probe is identical to that used in Table 1. **b**, The *c-fos* probe is a 1.3-kb *NaeI* fragment from rat *c-fos* cDNA³⁴. **c**, The *gag*-specific probe is a 0.8-kb *SstI*-*NcoI* fragment from pIV2.3T03, a Fujinami sarcoma virus derivative

(gift from M. Moran). The leucine zipper-DNA binding domain probe is a 279-bp fragment derived from the C terminus of *v-jun*, encompassing nucleotides 1,269–1,548 (ref. 1).

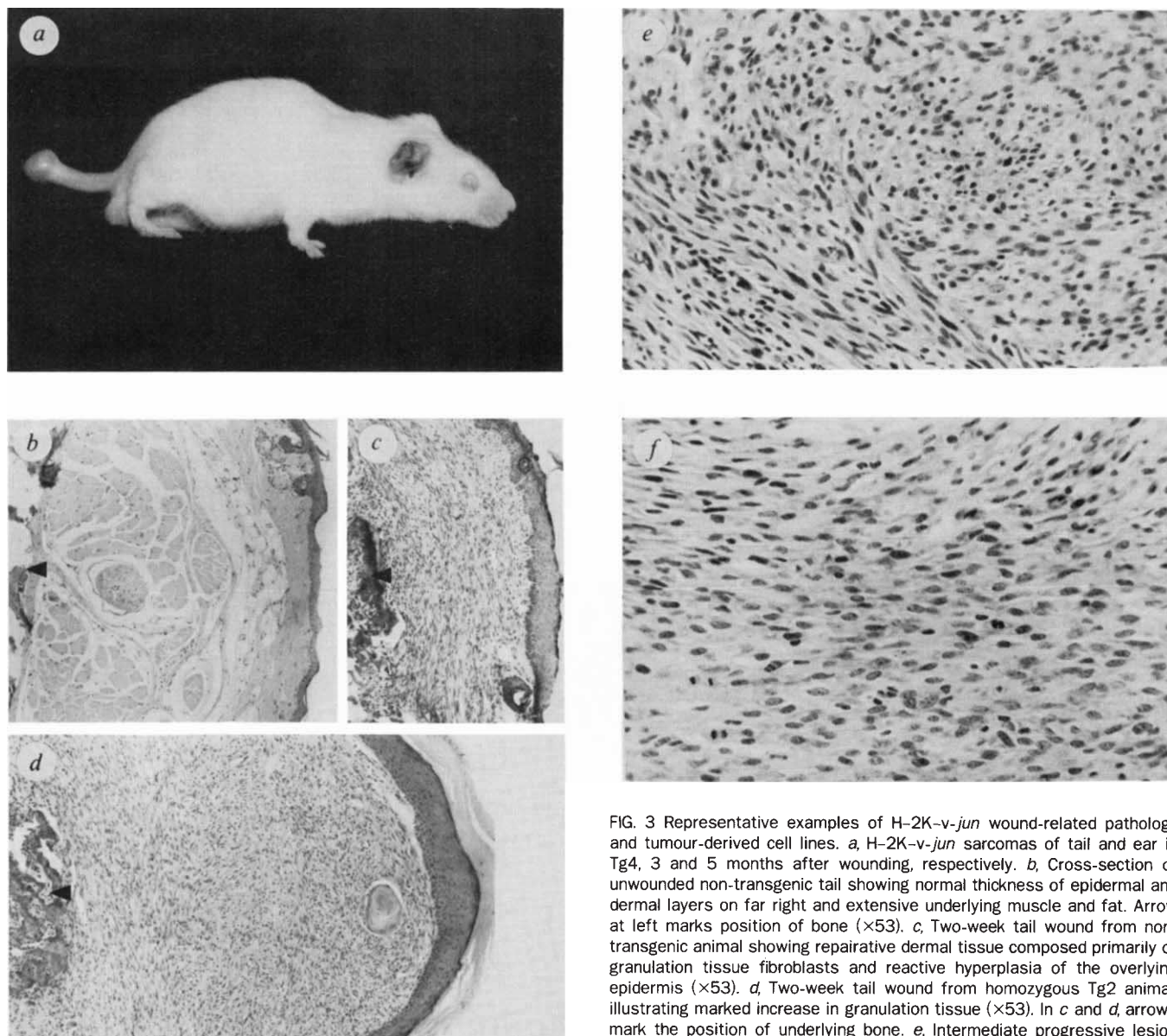
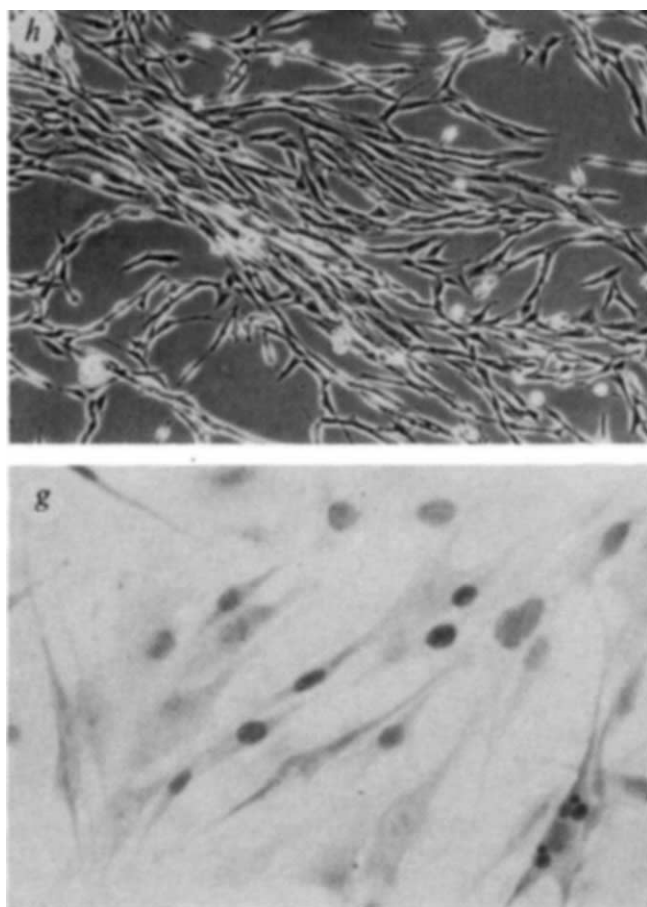


FIG. 3 Representative examples of H-2K-v-jun wound-related pathology and tumour-derived cell lines. *a*, H-2K-v-jun sarcomas of tail and ear in Tg4, 3 and 5 months after wounding, respectively. *b*, Cross-section of unwounded non-transgenic tail showing normal thickness of epidermal and dermal layers on far right and extensive underlying muscle and fat. Arrow at left marks position of bone ($\times 53$). *c*, Two-week tail wound from non-transgenic animal showing repairative dermal tissue composed primarily of granulation tissue fibroblasts and reactive hyperplasia of the overlying epidermis ($\times 53$). *d*, Two-week tail wound from homozygous Tg2 animal, illustrating marked increase in granulation tissue ($\times 53$). In *c* and *d*, arrows mark the position of underlying bone. *e*, Intermediate progressive lesion from homozygous Tg2 animal 10 weeks after wounding, showing extensive but benign fibroblastic proliferation ($\times 170$). *f*, Well-differentiated dermal fibrosarcoma from homozygous Tg2 animal, 15 weeks after tail wounding. Comparison with *e* reveals increased cellularity and mitotic index, increased nuclear pleomorphism, prominence of nuclear membranes and of nucleoli, and increased coarseness of chromatin ($\times 170$). *g*, Morphology of cell line from Tg4 tail fibrosarcoma (phase contrast, $\times 113$). *h*, Nuclear localization of v-jun by immunocytochemistry in cells derived from a Tg2 tail fibrosarcoma. Heterogeneous staining is consistently observed in clonally-derived

normal, a proportion of animals exhibits abnormal wound repair consisting of circumferential (and occasionally more focal) wound hypertrophy, following full-thickness wounding such as that which results from tail-cutting or ear-punching. This abnormality is usually unmistakable by 10–14 days following injury, and consists microscopically of hyperplastic granulation tissue (Fig. 3*d*) in large excess of that found in an identically produced, but non-transgenic wound (Fig. 3*c*). In many cases, abnormal growth ceases when the tissue defect is completely repaired but in others, the lesions continue to grow slowly for months, during which histological analysis reveals a benign but progressive dermal fibroblastic proliferation (Fig. 3*e*). Ultimately, a proportion of these intermediate lesions increases in size much more rapidly and becomes centrally haemorrhagic (Fig. 3*a*), and microscopic analysis is consistent with a diagnosis of well-differentiated fibrosarcoma (Fig. 3*f*). These malignancies can be locally invasive, but have not been observed to metastasize. The location and extent of wounding play a critical part in eliciting this phenotype. Abnormal wound repair and its subsequent progression have been observed primarily following wounds that completely or partially transect the ear or tail, but never after superficial epithelial injuries, suggesting that dermal (and possibly subdermal) injury or larger tissue defects that can be

repaired only by extensive fibroblast proliferation are particularly favoured for this outcome.

Tumorigenesis following wounding has been observed in all three transgenic lines, but with marked differences in frequency. In the Tg4 founder, the progression from abnormal tail wound to sarcoma occurred repeatedly after sequential tail cutting. In the Tg1 line, abnormal wound healing is observed in <5% of cases and so far, only one of these wounds has progressed to sarcoma. In Tg2, however, the entire sequence is observed frequently and predictably. Moreover, within this pedigree the wounding phenomenon and its subsequent progression are related to transgene dosage: After at least 3 months follow-up, in 50 hemizygous animals, 26 abnormal wounds (52%) gave rise to 10 slowly progressive lesions (38%), among which 4 sarcomas have been observed to occur at a mean of 20 weeks after



tumour cell populations.

METHODS. *b–f*, Tissue samples were fixed in 10% formaldehyde and then processed routinely for standard haematoxylin–eosin staining of 5- μ m paraffin sections. *g*, Sarcoma samples were minced finely and plated in RPMI 1640 or DME-high glucose medium supplemented with 10% FCS, 2 mM glutamine, 100 U ml⁻¹ penicillin, 100 μ g ml⁻¹ streptomycin, and 0.25 μ g ml⁻¹ amphotericin B. Medium was replaced twice weekly, and after 3 weeks, adherent cells extending from the tumour fragments were disaggregated with 0.05% trypsin/0.53 mM EDTA and replated. Clonal lines were subsequently established by limiting dilution. *h*, Cells were seeded on glass coverslips and incubated overnight in DME-high glucose medium supplemented as above, fixed in 4% paraformaldehyde in 10 mM PBS (10 min) and permeabilized in -20 °C acetone for 3 min. They were then rinsed in 50 mM Tris (pH 8.0), 150 mM NaCl and stained immunocytochemically using an avidin–biotinylated horseradish peroxidase kit (Vectastain ABC) with 0.05% diaminobenzidine tetrahydrochloride as peroxidase substrate and monoclonal anti-*gag* antibody, R254E (gift from T. Pawson) as primary antibody. This antibody detects the *gag* portion of v-Jun which is absent from cellular Jun proteins. Coverslips were then counterstained lightly with eosin.

wounding. By contrast, in 27 homozygous animals, 24 abnormal wounds (87%) gave rise to 18 intermediate lesions (75%), among which 6 sarcomas have been observed, occurring at a mean of 14 weeks following wounding. Thus, not only are abnormal wound repair and its progression more frequent in homozygous animals, but progression to sarcoma also occurs more rapidly.

Reproducible differences in transgene expression are observed in parallel with the morphological and corresponding histological progression from abnormal wound to sarcoma (Fig. 2*b*): *v-jun* mRNA is detectable at low levels in normal tail shaft, as well as in other tissues that may be found in tail such as skin, muscle, bone, and fibroblasts (Table 1). In 10–14 day-old hypertrophic wounds, there is roughly a twofold increase in *v-jun* mRNA, which persists in intermediate lesions 2–3 months old (which by this time may be ≥ 0.5 cm in diameter). The sub-

sequent conversion to sarcoma, however, is accompanied by an additional 5–10-fold increase in *v-jun* mRNA. The progression from abnormal wound to sarcoma, is not accompanied by any detectable changes in *c-fos* mRNA (Fig. 2*b*), nor could we demonstrate any accompanying changes in *c-jun* mRNA levels.

Several clonal cell lines have been established from Tg2 and Tg4 tail sarcomas (Fig. 3*g*). These cells are pleomorphic in appearance and include a component that can spontaneously give rise to multinucleated structures resembling myotubes (not shown). In addition, these cells form colonies in soft agar, are tumorigenic in nude mice (not shown), and can readily be shown to express *v-jun* by northern RNA analysis and by immunocytochemistry (Fig. 3*h*).

Our observations on tumour development and progression in H-2K-*v-jun* transgenic mice may provide some clues regarding the oncogenicity of *v-jun*. First, as only tail and ear fibrosarcomas have been observed, even though *v-jun* is expressed in a variety of tissues, all cell types may not be equally susceptible to transformation by *v-jun*, at least at the levels of expression attained. Second, *v-jun*-induced hyperplasia seems to depend on a transient wound-derived component. Wounds are known to mobilize a variety of growth factors and cytokines^{16–21} that may be interacting with v-Jun protein or with the transgene itself^{22–24} to induce proliferation. Third, the obligatory requirement for tissue injury and repair to initiate tumorigenesis indicates that even in tissues susceptible to *v-jun*, *v-jun* is not oncogenic as a 'single hit.' Indeed, in addition to the initiating wound, tumorigenesis probably requires early and late secondary genetic or epigenetic events that initially allow autonomous rather than wound-regulated proliferation, and later lead to increased *v-jun* expression. Fourth, the striking difference in wounding response between Tg1 and Tg2, despite similar *v-jun* mRNA levels in skin, muscle and bone, suggests that some other pedigree-specific feature of *v-jun* expression, such as inducibility within a wound or a subtle difference in cellular range of transgene expression, may be critical for tumorigenesis. Finally, and most intriguing, are the high levels of *v-jun* expression in sarcomas which imply distinct quantitative requirements or thresholds for *v-jun* action in early and late lesions.

The obligate relationship between wounding and appearance of tumours in H-2K-*v-jun* transgenic mice is consistent with anecdotal reports of this phenomenon in humans^{25–26} and with previous studies in birds²⁷ and mammals²⁸. Wound-associated mesenchymal neoplasms are also seen in transgenic mice carrying the bovine papilloma virus type-1 genome^{29–31}. The long latencies of tumour formation in these mice, however, suggest that these neoplasms, unlike those of H-2K-*v-jun* mice, have not arisen directly as an immediate outgrowth of the wound itself. The reproducibility of the wounding phenomenon in the H-2K-*v-jun* mice provides a useful model for studying tumour development and progression. Not only can the multistep conversion of an abnormality of a benign physiological process to frank malignancy be easily and accessibly studied, but the ability to isolate and analyse tissue from discrete stages in this sequence may allow identification of changes that lead to tumour progression, and how they may relate to normal components of tissue repair. Haddow proposed that "The wound may be regarded as a tumour which heals itself"²⁵; the *v-jun*-induced neoplasms that we describe reinforce this view by illustrating its corollary—the tumour as a wound that never heals. □

Received 5 March; accepted 24 May 1990.

1. Maki, Y., Bos, T. J., Davis, C., Starbuck, M. & Vogt, P. K. *Proc. natn. Acad. Sci. U.S.A.* **84**, 2848–2852 (1987).
2. Vogt, P. K. & Tjian, R. *Oncogene* **3**, 3–7 (1988).
3. Bos, T. J., Bohmann, D., Tsuchie, H., Tjian, R. & Vogt, P. K. *Cell* **52**, 705–712 (1988).
4. Bohmann, D. *et al. Science* **238**, 1386–1392 (1987).
5. Rauscher, F. J. III, *et al. Science* **240**, 1010–1016 (1988).
6. Angel, P. *et al. Nature* **332**, 166–171 (1988).
7. Ryder, K., Lau, L. F. & Nathans, D. *Proc. natn. Acad. Sci. U.S.A.* **85**, 1487–1491 (1988).
8. Ryder, K., Lanahan, A., Perez-Albuern, E. & Nathans, D. *Proc. natn. Acad. Sci. U.S.A.* **86**, 1500–1503 (1989).

9. Hirai, S.-I., Ryseck, R.-P., Mechta, F., Bravo, R. & Yaniv, M. *EMBO J.* **8**, 1433-1439 (1989).
10. Chiu, R. et al. *Cell* **54**, 541-552 (1988).
11. Halazonetis, T. D., Georgopoulos, K., Greenberg, M. E. & Leder, P. *Cell* **55**, 917-924 (1988).
12. Lee, W., Haslinger, A., Karin, M. & Tjian, R. *Nature* **325**, 368-372 (1987).
13. Lee, W., Mitchell, P. & Tjian, R. *Cell* **49**, 741-752 (1987).
14. Angel, P. et al. *Cell* **49**, 729-739 (1987).
15. Arnold, B., Burgert, H.-G., Archibald, A. L. & Kvist, S. *Nucleic Acids Res.* **12**, 9473-9487 (1984).
16. Clark, R. A. F. in *The Molecular and Cellular Biology of Wound Repair* (eds Clark, R. A. F. & Henson, P. M.) 3-33 (Plenum, New York, 1988).
17. Sporn, M. B. & Roberts, A. B. *J. clin. Invest.* **78**, 329-332 (1986).
18. Brenner, M. K. *Br. J. Haemat.* **69**, 149-152 (1988).
19. Steenfos, H. H., Hunt, T. K., Scheuenstuhl, H. & Goodson, W. H. *Surgery* **106**, 171-176 (1989).
20. Dinarello, C. A. *FASEB J.* **2**, 108-115 (1988).
21. Rappolee, D. A., Mark, D., Banda, M. J. & Werb, Z. *Science* **241**, 708-712 (1988).
22. Israel, A., Kimura, A., Fournier, A., Fellous, M. & Kourilsky, P. *Nature* **322**, 743-746 (1986).
23. Israel, A. et al. *EMBO J.* **8**, 3793-3800 (1989).
24. Muegge, K. et al. *Science* **246**, 249-251 (1989).
25. Haddow, A. *Adv. Cancer Res.* **16**, 181-234 (1972).
26. Van Den Hooff, A. *Adv. Cancer Res.* **50**, 159-195 (1988).
27. Dolberg, D. S., Hollingsworth, R., Hertle, M. & Bissell, M. J. *Science* **230**, 676-678 (1985).
28. Argyris, T. S. *CRC Crit. Rev. Toxicol.* **14**, 211-258 (1984).
29. Lacey, M., Alpert, S. & Hanahan, D. *Nature* **322**, 609-612 (1986).
30. Hanahan, D., Alpert, S. & Lacey, M. in *Viral Carcinogenesis* (eds Kjeldgaard, N.O. & Forchhammer, J.) 263-274 (Munksgaard, Copenhagen, 1987).
31. Sippola-Thiele, M., Hanahan, D. & Howley, P. M. *Molec. cell. Biol.* **9**, 925-934 (1989).
32. Breitman, M. L. et al. *Science* **238**, 1563-1565 (1987).
33. Hogan, B., Constantini, F. & Lacy, E. (eds). *Manipulating the Mouse Embryo: A Laboratory Manual* (Cold Spring Harbor Laboratory, 1986).
34. Curran, T., Gordon, M. B., Rubino, K. L. & Sambucetti, L. C. *Oncogene* **2**, 79-84 (1987).

ACKNOWLEDGEMENTS. We thank Anthony Zucconi and Todd MacGregor for technical assistance, and Frances Hogue and Lynda Woodcock for typing the manuscript. This work was supported by grants from the NCI (Canada; M.L.B.) and from the NCI (US) and the Council for Tobacco Research (P.K.V.). M.L.B. is an MRC Scientist, and A.C.S. is the recipient of an MRC Centennial Fellowship.

Human cyclin A is adenovirus E1A-associated protein p60 and behaves differently from cyclin B

Jonathon Pines & Tony Hunter*

Molecular Biology and Virology Laboratory,
The Salk Institute for Biological Studies, PO Box 85800, San Diego,
California 92138, USA

CYCLINS are proteins synthesized during each cell cycle and abruptly destroyed in each mitosis¹⁻⁵. Cyclins have been implicated in the induction of mitosis^{6,7} and are associated with the serine-threonine protein kinase p34^{cdc2} as components of mitosis promoting factor (MPF)⁸⁻¹⁰. On the basis of conserved sequence motifs cyclins can be divided into A or B types. We recently cloned a human cyclin B (ref. 11) and showed that cyclin B expression is regulated transcriptionally and post-translationally during the cell cycle, and that cyclin B associates with p34^{cdc2}. Here we report that human cyclin A messenger RNA and protein levels also vary during the cell cycle, and increase and decrease in advance of cyclin B levels. Cyclin A is associated with a protein of relative molecular mass 33,000 that is related to, but distinct from, p34^{cdc2}, and this complex has histone H1 kinase activity *in vitro*. Cyclin A is identical to p60, a protein that associates with p34^{cdc2} in interphase cells¹² and with adenovirus E1A in transformed cells²².

We isolated a human cyclin A complementary DNA using the polymerase chain reaction (PCR). Degenerate oligonucleotides to two conserved cyclin A motifs were designed on the basis of the known cyclin A sequences from *Drosophila*¹³, *Spisula*² and *Xenopus* (T. Hunt, personal communication). As template we used the HeLa cell cDNA library¹⁴ from which we had isolated a cyclin B cDNA clone. Using the cloned PCR product as a probe, we isolated a 1.4-kilobase (kb) cDNA and in turn used this clone to isolate a 2.2-kb cDNA from a human keratinocyte cDNA library. Recently a human cyclin A cDNA has been independently cloned¹⁵ and the amino-acid sequence reported is identical to our own. We detected a single human cyclin A mRNA of about 2.2 kb; it is not clear why this is apparently smaller than the main 2.7-kb cyclin A cDNA previously reported¹⁵.

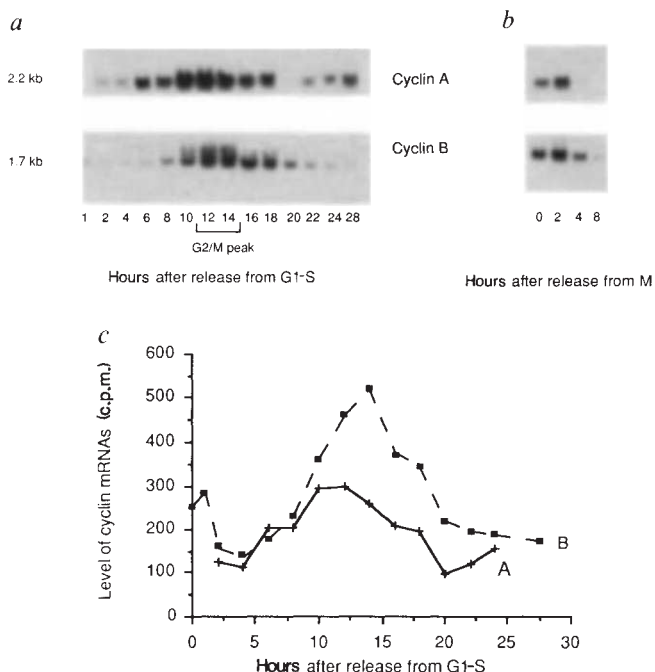


FIG. 1 Cyclin A and B mRNA levels vary during the cell cycle. *a*, Northern blot of RNA from cells after release from a G1-S phase block, probed with cyclin A and B cDNAs. Cells were synchronized by a thymidine-aphidicolin block and RNA was prepared at the indicated times as described¹¹. The blots were probed with M13 sequencing clones of either cyclin A or B prepared by primed-cut labelling. Flow cytometry showed that the M-phase peak was at 12-14 h. RNA sizes are given in kb on the left. *b*, Northern blot of RNA from cells after release from a mitotic block. RNA was prepared and blotted as above, at the indicated times, from cells released from a nocodazole block¹¹. The apparent increase in cyclin mRNAs after the release from nocodazole block is a loading artefact. When the cyclin mRNA levels are normalized to actin mRNA they remain constant for 2 h after removal of nocodazole and then begin to decline. *c*, Relative intensity of cyclin A and B mRNA signals from the blots shown in *a* as measured by a beta scanner and normalized to that of actin¹¹.

METHODS. RNA was prepared from $\sim 8 \times 10^6$ HeLa cells per time point. The same RNA samples were used for the cyclin A and B blots in *a* and *b*. Total RNA was loaded (10 μ g per lane) on a 1% agarose-formaldehyde gel and the RNA transferred and probed as described¹¹. The cyclin A blot was exposed with an intensifying screen for 3 h at -70°C , and the cyclin B blot for 16 h at room temperature without a screen. RNA markers (BRL) were run in parallel.

We find that the level of cyclin A mRNA is regulated in the cell cycle. In G2 there is roughly three times more cyclin A mRNA than in G1, and this reaches its maximal level 2 h before the peak in the amount of cyclin B mRNA (Fig. 1*a*). The rate of transcription of cyclin A mRNA also increases about 2 h before that of cyclin B as determined by nuclear run-off experiments (data not shown). In addition to an increased rate of transcription, cyclin A mRNA may be more stable in G2 than in G1. The level of cyclin A mRNA declines more rapidly after mitosis than that of cyclin B (Fig. 1*a, b*), implying that cyclin A mRNA is more unstable than cyclin B mRNA in G1, consistent with the presence of several AUUUA motifs in its 3' untranslated region, which are not found in cyclin B mRNA¹¹.

The cyclin A *in vitro* translation product has an apparent relative molecular mass (M_r) of 58,000 (58K), as determined by SDS-PAGE (data not shown). This, coupled with the histone H1 kinase activity of cyclin A immunoprecipitates (see below), indicated that cyclin A could be related to the p60 protein that associates with p34^{cdc2} (ref. 12). In support of this we found that the cyclin A *in vitro* translation product was immunoprecipitated with the anti-p60 monoclonal antibody C160 (ref. 12). Furthermore, the [³⁵S]methionine-labelled tryptic-peptide map of the *in vitro* translation product of cyclin A is identical to that

* To whom correspondence should be addressed.

FIG. 2 Cyclin A is identical to p60. *a*, [^{35}S]methionine tryptic peptide map of cyclin A *in vitro* translation product (8,500 c.p.m. applied, exposed for 72 h). *b*, [^{35}S]methionine tryptic peptide map of p60 immunoprecipitated from HeLa cells with monoclonal antibody C160 (950 c.p.m. applied, exposed for 500 h). *c*, [^{35}S]methionine tryptic peptide map of mix of samples from *a* and *b* (800 c.p.m. of each sample applied, exposed for 500 h). METHODS. Cyclin A synthetic mRNA was transcribed and translated *in vitro* from an SP6 promoter as previously described¹¹. Roughly 2×10^6 HeLa cells were labelled for 2 h in methionine-free DMEM + 10% dialysed fetal bovine serum supplemented with 0.5 mCi ml⁻¹ [^{35}S]methionine (Amersham). Cells were lysed in 1 ml RIPA buffer and 10 μl monoclonal antibody C160 cell supernatant used to immunoprecipitate p60 per 100 μl lysate. Tryptic peptides were prepared and run as two-dimensional maps by electrophoresis at pH 1.9 in the horizontal direction (anode on the left), and by ascending chromatography¹¹. The origin is marked by an arrow.

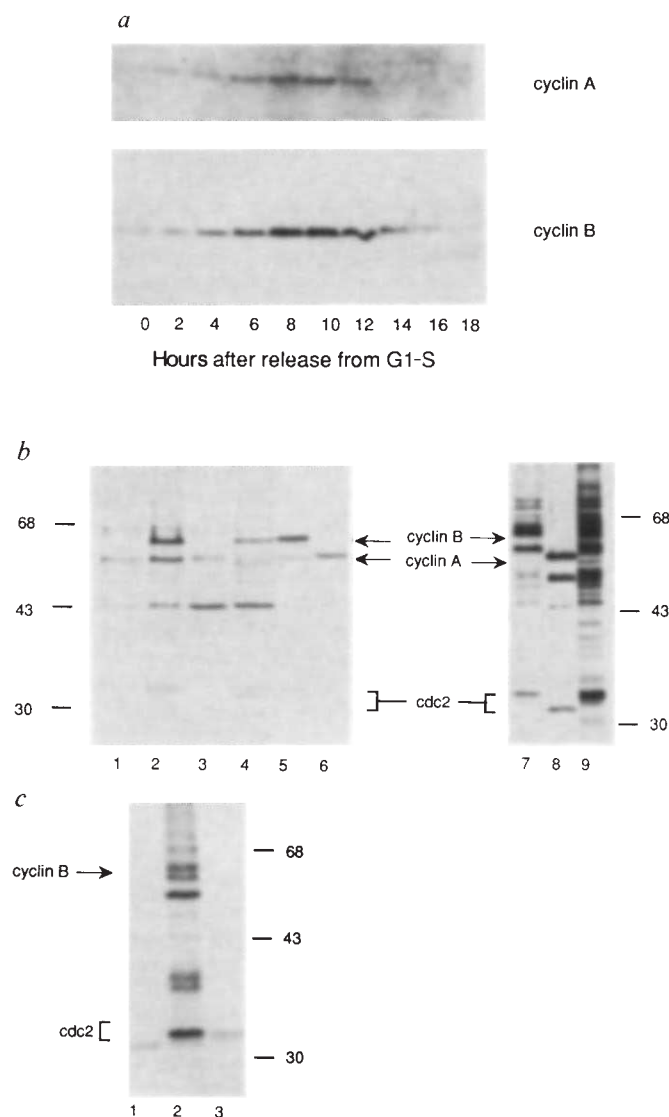
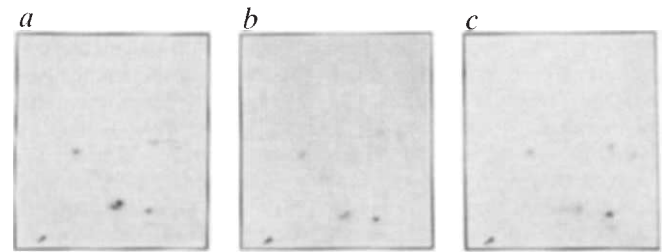


FIG. 3 Cyclin A and B protein levels alter during the cell cycle and co-immunoprecipitate with different ~34K proteins. *a*, Immunoblot. HeLa cells were released from an aphidicolin block and total cell lysates were prepared in SDS lysis buffer at the indicated times. Equal amounts of protein were loaded per lane. The blot was probed with either anti-cyclin A or anti-cyclin B antiserum followed by ^{125}I -labelled protein A as previously described¹¹. Immunoblots were exposed at -70°C with an intensifying screen for 150 h (cyclin A) or 15 h (cyclin B). Flow cytometry showed the G2-M peak to be at 14 h. *b*, Immunoprecipitations from [^{35}S]methionine/cysteine-labelled HeLa cells. Lane 1: cyclin B, preimmune serum; lane 2: anti-cyclin B (exposed for 12 h); lane 3: anti-cyclin A (exposed for 14 h); lane 4: anti-cdc2 C-terminal peptide (exposed for 14 h); lane 5: p13^{suc1}-Sepharose eluate (exposed for

14 h); lane 6: anti-cyclin A (exposed for 14 h); lane 7: anti-cyclin B (exposed 12 h); lane 8: anti-cyclin A (exposed 5 days); lane 9: longer exposure (5 days) of lane 7. *M_r* markers (in K) are on the left and right. Lanes 1–6, and lanes 7–9 were run on separate gels. Lanes 4 and 9 show the positions of all three forms of p34^{cdc2}. *c*, Immunoprecipitation of cyclin A (lane 1), cyclin B (lane 2) and p34^{cdc2} (lane 3) from ^{32}P -labelled HeLa cells. The gel was exposed for 15 h with an intensifying screen at room temperature. *M_r* markers (in K) are on the right of the figure. *d*, Re-immunoprecipitation of the 33–34K proteins from ^{35}S -labelled anti-cyclin A or anti-cyclin B immunoprecipitates with either an anti-PSTAIR or an anti-p34^{cdc2} C-terminal peptide antibody. The exposure shown was for 40 h, but much longer exposures confirm that the cyclin A-associated 33K protein is not re-immunoprecipitated with the C-terminal antibody.

METHODS. For immunoblots $\sim 0.5 \times 10^6$ cells were lysed in 100 μl SDS lysis buffer per time point. The concentration of protein in each sample was measured by the BCA protein assay (Pierce Chemicals) and 33 μg of each sample was loaded per lane. Polyclonal anti-cyclin A or anti-cyclin B antisera were raised in rabbits to a bacterially expressed fusion protein of the TrpE protein fused to the C-terminal 46 amino acids of cyclin A. For immunoblots, anti-cyclin A antiserum was used at a final concentration of 1:250, anti-cyclin B antiserum was used at 1:1000. An extra wash in 0.1% Nonidet P-40, 150 mM NaCl, 10 mM Tris-HCl buffer, pH 7.4, removed the signal from the $\sim 58\text{K}$ protein previously recognized by anti-cyclin B antibodies¹¹. The signals obtained for cyclin A and B on immunoblots are not equivalent because the titres of the antisera differ. Asynchronously growing cells were labelled for 2 h with [^{35}S]methionine/cysteine or for 3 h with ^{32}P -orthophosphate under conditions previously described¹¹. Cells were lysed in RIPA buffer and cyclin B immunoprecipitations were prepared and analysed as previously described¹¹. For cyclin A immunoprecipitations, 10 μl monoclonal antibody C160 cell supernatant was used per 100 μg cell lysate. For p34^{cdc2} immunoprecipitates, 1 μl anti-p34^{cdc2} C-terminal peptide antiserum was used per 100 μg of cell lysate. The same ^{35}S -labelled lysates were used for all the immunoprecipitations in *b*, and the samples were run on the same gel. Re-immunoprecipitations of the 33–34K proteins from cyclin A or B immunocomplexes were carried out as described⁴.

of p60 immunoprecipitated with C160 (Fig. 2). We conclude that p60 is cyclin A.

Antibodies raised to a bacterially expressed cyclin A fusion protein immunoprecipitated a 60K protein that had an identical peptide map to p60 (data not shown). We used these antibodies and the C160 monoclonal antibody to study the level and activity of cyclin A through the cell cycle compared with cyclin B. By immunoblotting we found the amount of both cyclin A and B increases through the S and G2 phases to a maximum at mitosis, at which point they are degraded (Fig. 3a). Cyclin A destruction seems to be initiated before that of cyclin B (Fig. 3a), which is in keeping with the behaviour of the cyclins in clams, where cyclin A is degraded first in metaphase, and cyclin B later at the metaphase-anaphase transition^{4,16,17}.

Both cyclin A and cyclin B co-immunoprecipitate with proteins of around 34K from either ³⁵S- or ³²P-labelled cells (Fig. 3b and c). We have previously identified by tryptic peptide mapping the 34K protein associated with cyclin B as p34^{cdc2} (ref. 11). As expected, this 34K protein is recognized by both an anti-C-terminal p34^{cdc2} peptide antibody and by an antibody raised against the conserved PSTAIR motif (one-letter amino-acid notation) when released from an anti-cyclin B ³²P-labelled immunoprecipitate (Fig. 3d). By contrast, cyclin A seems to be mostly associated with a ~33K protein related to, but distinct from p34^{cdc2}. The ~33K protein associated with cyclin A migrates slightly ahead of all three forms of p34^{cdc2} immunoprecipitated with an anti-C-terminal p34^{cdc2} peptide antibody (Fig. 3b, lanes 4, 8 and 9). The faint band just above the 33K protein in the anti-cyclin A immunoprecipitate in lane 8 could be p34^{cdc2} as it comigrates with the fastest migrating (least phosphorylated) form of p34^{cdc2}, but there was too little material to analyse further. The 33K protein is strongly labelled with ³²P *in vivo* (Fig. 3c), arguing against its being p34^{cdc2} because the highly phosphorylated form of p34^{cdc2} migrates at a position corresponding to ~35K. When the 33K protein is released from a ³⁵S-labelled anti-cyclin A immunoprecipitate it can be reprecipitated with an anti-PSTAIR peptide antibody, but not with an anti-C-terminal p34^{cdc2} peptide antibody (Fig. 3d). Tryptic peptide mapping confirms that this protein is distinct from the p34^{cdc2} detected in cyclin B immunoprecipitates (data not shown). Consistent with our conclusion that cyclin A does not primarily associate with p34^{cdc2}, we do not detect a 60K protein comigrating with cyclin A in immune complexes made with the anti-C-terminal p34^{cdc2} peptide antibody (Fig.

3b, lane 4). Interestingly, we do detect a 60K protein with an identical tryptic peptide map to cyclin A bound to p13^{suc1}-Sephadex (Fig. 3b, lane 5), which suggests that p13^{suc1} binds both p34^{cdc2} and closely related proteins, such as the 33K protein.

Despite the apparent absence of p34^{cdc2} in cyclin A immunoprecipitates, these immunocomplexes have histone H1 kinase activity *in vitro* (Fig. 4), but differ from those of cyclin B. Cyclin B-associated H1 kinase activity increases dramatically as the cells enter mitosis. In contrast, the cyclin A-associated H1 kinase activity increases only slightly and gradually through S and G2 phase, roughly paralleling the rise in cyclin A protein, and reaches a maximum before that of cyclin B (Fig. 4). The cyclin A-associated H1 kinase activity begins to decrease before that of cyclin B, consistent with the earlier decline in cyclin A protein. As in clams¹⁶, human cyclin B is stable in cells arrested with microtubule inhibitors such as nocodazole, but cyclin A is not (data not shown). This results in an artefactually low level of cyclin A-associated histone H1 kinase activity in nocodazole-arrested cells and explains why this complex was proposed to be the interphase form of p34^{cdc2} (ref. 12).

At first sight our results seem to conflict with the finding in clams that cyclin A associates with p34^{cdc2} (ref. 8). But in clams cyclin A is not reliably coprecipitated with anti-p34^{cdc2} antibodies, although it is easily detected using p13^{suc1}-Sephadex. In studies on HeLa cells Giordano *et al.*¹² found that p60 could be immunoprecipitated by anti-cdc2 C-terminal antibody, but the level of p60 was considerably less than that of cyclin B. So far we have been unable to detect cyclin A-associated p34^{cdc2} by immunoblotting the anti-cyclin A immunoprecipitates with an anti-C-terminal p34^{cdc2} peptide antibody (data not shown), but it remains possible that cyclin A binds to a minor fraction of the underphosphorylated form of p34^{cdc2}.

In summary, we have shown that human cyclin A mRNA and protein levels are regulated in the cell cycle and rise and fall in advance of those of cyclin B. The close but staggered induction of cyclin A and B mRNA transcription in the cell cycle shows that this system is tightly regulated at the nuclear level. There are several differences at the protein level between cyclin A and cyclin B, notably in the proteins with which they associate and the timing of their associated histone H1 kinase activities. Cyclin B associates with p34^{cdc2} to form a complex in G2 phase, but this does not become fully active as a protein kinase until the p34^{cdc2} subunit is dephosphorylated on phosphotyrosine and

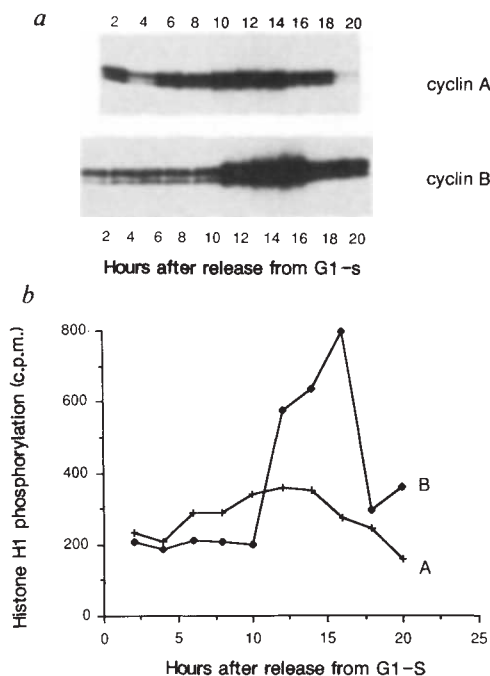


FIG. 4 *In vitro* kinase activities associated with cyclin A and B vary during the cell cycle. a, Kinase assay of cyclin A or B immunoprecipitates with histone H1 as a substrate. Cells were synchronized by a thymidine-aphidicolin block and samples taken at the indicated times after release from the block. The autoradiograph was exposed without a screen for 14 h at room temperature. Flow cytometry analysis showed the G2-M peak to be at 14 h. b, Incorporation of label into histone H1 from the gel in panel A estimated by a beta scanner.

METHODS. For kinase assays, 1×10^6 cells were lysed in RIPA buffer per time point and H1 kinase assays and cyclin B immunoprecipitations were prepared and performed as previously described¹¹. For cyclin A immunoprecipitations, 10 μ l monoclonal antibody C160 cell supernatant was used per 100 μ g cell lysate. The same lysates were used for the anti-cyclin A and the anti-cyclin B immunoprecipitations.

phosphothreonine in late G2 (refs 18, 19). In contrast, cyclin A associates with a 33K protein related to p34^{cdc2} and this complex does not show a dramatic increase in *in vitro* kinase activity as cells enter M phase, but rather its kinase activity gradually increases as cells progress from S to G2 phase. Although we have not demonstrated that the 33K protein found in anti-cyclin A immunoprecipitates is a protein kinase, this seems likely from its close relation to p34^{cdc2}; not only does the 33K protein contain the PSTAIR motif, but it also has a very similar fragmentation pattern to p34^{cdc2} when cleaved with *N*-chlorosuccinimide¹². (There are now several reports of cdc2-related proteins that have the PSTAIR motif in other species, notably *Xenopus*, *Drosophila* and *Saccaromyces cerevisiae*²⁰, which all have four conserved tryptophans, the residue recognized by *N*-chlorosuccinimide). In addition, cyclin A itself is not likely to be a protein kinase as its sequence does not have any of the conserved motifs found in all known protein kinases. So far few differences have been noted in the protein kinase specificities of the two types of cyclin complex *in vitro*²⁰, which also implies that the cyclin A-associated kinase is very similar to p34^{cdc2}. Like the form of p34^{cdc2} associated with cyclin B, the 33K protein associated with cyclin A contains phosphotyrosine (data not shown). Thus, by analogy with p34^{cdc2}, the protein kinase activity of the cyclin A complex might also be regulated by phosphorylation of its catalytic subunit.

As cyclin A complexes have protein kinase activity earlier in the cell cycle than the cyclin B-p34^{cdc2} complex, it is possible that the cyclin A complex is involved in the S phase itself, fulfilling the role in human cells performed by the 'G1 cyclins' in *S. cerevisiae*²¹. However, this seems unlikely as we find that human cyclin A mRNA causes *Xenopus* oocyte maturation, which implicates it in the induction of mitosis.

It is intriguing that cyclin A has now been indirectly linked with cellular transformation in two ways. First, p60 was initially described through its association with E1A in adenovirus-transformed cells²². E1A binds to and probably inactivates the retinoblastoma protein^{23,24}, and in an analogous fashion may also modulate the function of cyclin A. Second, hepatitis B virus sequences were found integrated into an intron of the cyclin A gene in a hepatocellular carcinoma¹⁵. Thus it is possible that disrupting the normal function of cyclin A has profound effects on the regulation of the cell cycle. The nature of this disruption and whether the same is also true for cyclin B now need to be investigated. □

Received 16 March; accepted 13 June 1990.

- Evans, T., Rosenthal, E. T., Youngblom, J., Distel, D. & Hunt, T. *Cell* **33**, 389–396 (1983).
- Swenson, K., Farrell, K. M. & Ruderman, J. V. *Cell* **47**, 861–870 (1986).
- Standart, N., Minshall, J., Pines, J. & Hunt, T. *Dev Biol.* **124**, 248–258 (1987).
- Westendorf, J. M., Swenson, K. I. & Ruderman, J. V. *J. Cell Biol.* **108**, 1431–1444 (1989).
- Hunt, T. *Curr. Opin. Cell Biol.* **1**, 286–274 (1989).
- Minshall, J., Blow, J. J. & Hunt, T. *Cell* **56**, 947–956 (1989).
- Murray, A. W. & Kirschner, M. W. *Nature* **339**, 275–280 (1989).
- Draetta, G. et al. *Cell* **56**, 829–838 (1989).
- Labbé, J.-C. et al. *EMBO J.* **8**, 3053–3058 (1989).
- Gautier, J. et al. *Cell* **60**, 487–494 (1990).
- Pines, J. & Hunter, T. *Cell* **58**, 833–846 (1989).
- Giordano, A. et al. *Cell* **58**, 981–990 (1989).
- Lehner, C. F. & O'Farrell, P. H. *Cell* **56**, 957–968 (1989).
- Hanks, S. K. *Proc. natn. Acad. Sci. U.S.A.* **84**, 388–392 (1987).
- Wang, J., Chenivresse, X., Henglein, B. & Bréchet, C. *Nature* **343**, 555–557 (1990).
- Minshall, J. et al. in *Cell Cycle Control in Eukaryotes* (eds Beach, D., Basilio, C. & Newport, J.) 128–139 (Cold Spring Harbor Laboratory, New York, 1988).
- Luca, F. C. & Ruderman, J. V. *J. Cell Biol.* **109**, 1895–1909 (1989).
- Gould, K. L. & Nurse, P. *Nature* **342**, 39–45 (1989).
- Pondaven, P., Meijer, L. & Beach, D. *Genes Dev.* **4**, 9–17 (1990).
- Pines, J. & Hunter, T. *New Biologist* **2**, 389–401 (1990).
- Richardson, H. E., Wittenberg, C., Cross, F. & Reed, S. I. *Cell* **59**, 1127–1133 (1989).
- Harlow, E., Whyte, P., Franza, B. R., Jr. & Schley, C. *Molec. cell. Biol.* **6**, 1579–1589 (1986).
- Whyte, P. et al. *Nature* **334**, 124–129 (1988).
- De Caprio, J. A. et al. *Cell* **58**, 1085–1095 (1989).

ACKNOWLEDGEMENTS. We thank W. Boyle for pointing out that p60 could be cyclin A and other insightful comments, G. Draetta and E. Harlow for C160 monoclonal antibody, D. Beach for anti-cyclin A antiserum, P. Russell for anti-PSTAIR antiserum, R. Butcher and D. Deguzman for oligonucleotide synthesis, D. Deguzman for flow cytometry, J. Kanki for oocyte microinjections, R. Lindberg for the keratinocyte cDNA library, S. Hanks for constructive criticism and T. Hunt and M. Dasso for perceptive and stimulating discussions. This work was supported by the NCI and the American Business Foundation for Cancer Research. J.P. was the recipient of an EMBO Long-term Fellowship.

Sequential activation of *HOX2* homeobox genes by retinoic acid in human embryonal carcinoma cells

Antonio Simeone*, Dario Acampora*, Laura Arcioni†, Peter W. Andrews‡, Edoardo Boncinelli* & Fulvio Mavilio†§

* International Institute of Genetics and Biophysics, CNR, Via Marconi 10, 80125 Napoli, Italy

† Istituto Scientifico H. S. Raffaele, Via Olgettina 60, 20132 Milano, Italy

‡ The Wistar Institute, 36th and Spruce Streets, Philadelphia, Pennsylvania 19104, USA

RETINOIC acid has been implicated as a natural morphogen in chicken^{1–4} and frog⁵ embryogenesis, and is presumed to act through the gene regulatory activity of a family of nuclear receptors^{6–10}. Homeobox genes, which specify positional information in *Drosophila* and possibly in vertebrate embryogenesis^{11–21}, are among the candidate responsive genes^{4,5,14}. We previously reported that retinoic acid specifically induces human homeobox gene (*HOX*) expression in the embryonal carcinoma cell line NT2/D1 (ref. 22). We now show that the nine genes of the *HOX2* cluster are differentially activated in NT2/D1 cells exposed to retinoic acid concentrations ranging from 10^{–8} to 10^{–5} M. Genes located in the 3' half of the cluster are induced at peak levels by 10^{–8} M retinoic acid, whereas a concentration of 10^{–6} to 10^{–5} M is required to fully activate 5' genes. At both high and low retinoic acid concentrations, *HOX2* genes are sequentially activated in embryonal carcinoma cells in the 3' to 5' direction.

The cloned human embryonal carcinoma (EC) cell line NT2/D1 shows the phenotype of multipotent embryonic stem cells, and is induced to differentiate by retinoic acid (RA) into a variety of cell types, including neurons²³. Cells were cultured for 14 days in medium containing RA at different concentrations in the non-toxic range of 10^{–8} to 10^{–5} M. Differentiation was monitored by the proportion of cells positive to glycolipid surface markers expressed in either stem or differentiated EC cells²³. An RA concentration of 5 × 10^{–8} M was sufficient to cause disappearance of ~50% stem cells and significant expression of differentiation antigens (Fig. 1a). Total RNA was extracted from 14-day cultures and probed for expression of *HOX2* genes. Maximal levels of transcript accumulation from genes located in the 5' half of the *HOX2* cluster (that is, from *HOX2E* to *HOX2A*) occurred in cells exposed to at least 5 × 10^{–6} M RA, whereas decreasing RNA levels were detected in cells exposed to lower RA concentrations (Fig. 1b). Conversely, transcripts from the 3' genes *HOX2F*, *HOX2H* and *HOX2I* were accumulated at peak levels at RA concentrations of around 10^{–8} M. Thus, *HOX2* genes are differentially activated by RA in EC cells, that is, 3' genes are activated at maximum levels by RA concentrations that are two orders of magnitude lower than those required to fully activate 5' genes. *HOX2A* apparently marks the border between the two groups. An apparent exception is represented by *HOX2G*, which despite being located in the 3' group showed an activation pattern more typical of the 5' group (Fig. 1b).

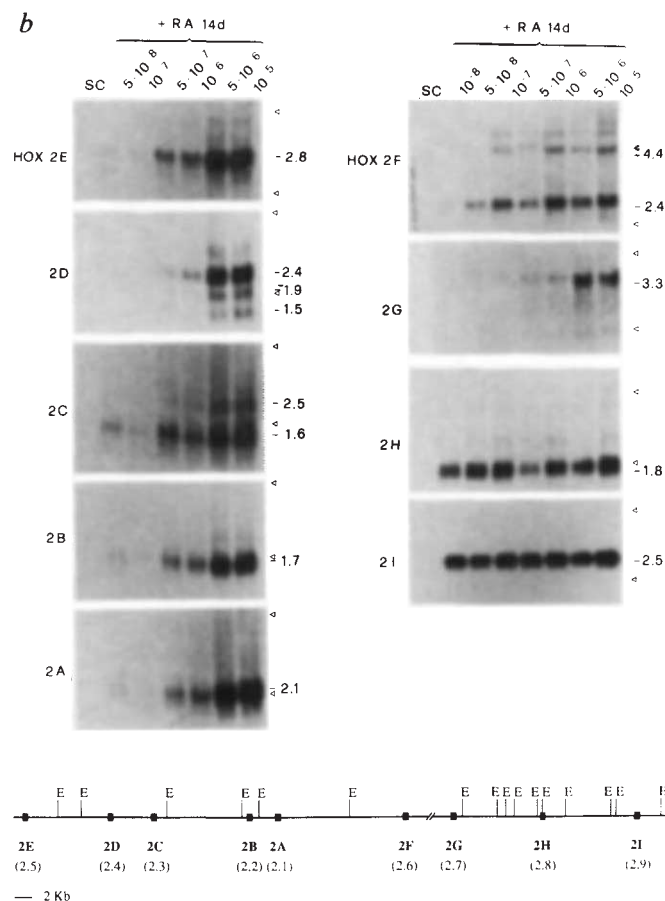
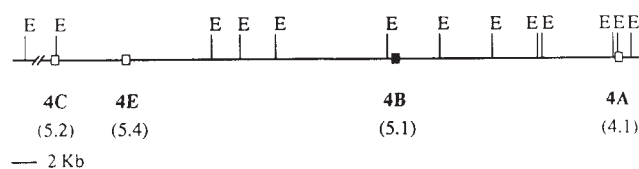
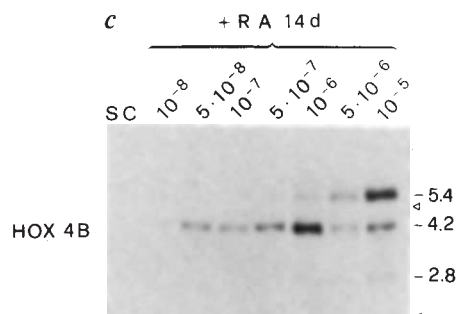
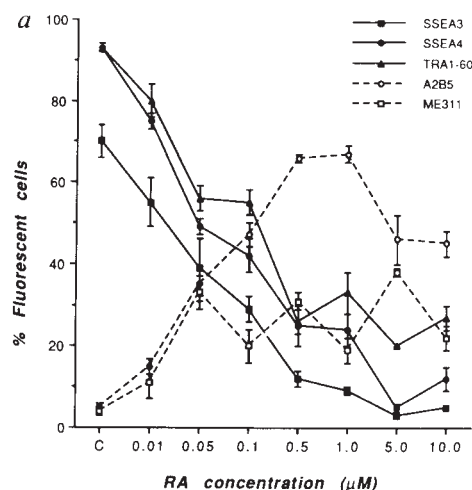
The number and size of *HOX2* gene transcripts observed in RA-induced EC cells were virtually identical to those observed in human embryos^{15,17,20}. We also analysed the expression pattern of *HOX4B*, which is expressed in human embryos in several transcripts according to a spatially restricted and stage-specific pattern¹⁶. Northern blot analysis indicated that a 4.2-kilobase (kb) transcript accumulated in NT2/D1 cells after exposure to

§ To whom correspondence should be addressed.

10^{-8} M RA, whereas 5.4-kb and 2.8-kb transcripts became increasingly abundant starting from 10^{-6} M RA (Fig. 1c). The anterior border of expression of the 4.2-kb transcript in 6–7-week human embryonic central nervous system (CNS) is in the hind-brain region, whereas 5.4- and 2.8-kb transcripts are expressed more caudally in the spinal cord, as well as in the limb bud¹⁶. The present data indicate that different RA concentrations direct alternative expression patterns from *HOX4B* in EC cells, reproducing a region-specific developmental regulation observed *in vivo*.

We then studied the time course of the RA-induced *HOX2* gene activation in conditions in which all *HOX2* genes, or only the 3' genes are activated, that is, in the presence of 10^{-5} M or

10^{-7} M RA respectively. Expression of single genes was analysed by RNase protection assay. At 10^{-5} M RA, transcripts from *HOX2H* and *HOX2I* rapidly and steadily accumulated, reaching peak levels at 24 h of RA exposure. Activation of the other genes proceeded sequentially from *HOX2G* to *HOX2E* between 1 and 4 days of RA induction (Fig. 2a). Fifty per cent maximal RNA accumulation levels were reached at 14 h for *HOX2I* and *HOX2H*, 60 h for *HOX2G*, *HOX2F* and *HOX2A*, 70 h for *HOX2B*, 90 h for *HOX2C*, 115 h for *HOX2D* and 180 h for *HOX2E* (Fig. 2b). The *HOX2* cluster is thus progressively activated in the 3' to 5' direction, according to a discrete pattern. As previously reported, continuous presence of RA is required throughout the induction process to activate 5' genes²².



HOX4 maps are 18 and 15 kb respectively.

METHODS. The cloned, human EC cell line NT2/D1 (ref. 23) was maintained in high-glucose DMEM supplemented with 10% FCS in a humidified atmosphere of 5% CO₂ in air. Differentiation was induced by seeding cells at 5×10^5 per 75 cm² flask in 10^{-8} – 10^{-5} M RA (all-*trans*, Sigma Chemical Co.). Cells were refed every 4 days with RA-containing medium and analysed after 14 days by indirect immunofluorescence, as previously described^{22,23}. Total RNA was extracted by the guanidine-isothiocyanate technique²⁷, selected for poly(A)⁺ by one passage on oligo(dT) cellulose columns according to standard procedures, run in 2–3 μ g aliquots on 1.0% agarose-formaldehyde gel, transferred by northern capillary blot²⁸ onto nylon membranes (Hybond-N, Amersham) and hybridized to 10^7 c.p.m. DNA probe labelled by random priming to a specific activity of 10^9 d.p.m. μ g⁻¹. Filters were washed under stringent conditions (0.1 $\times$ SSC buffer, 0.2% sodium dodecyl sulphate (SDS) at 65 °C for 60 min) and exposed to Kodak XAR-5 film at –70 °C for 1–3 days. Filters were stripped and reprobed up to six times as previously described²². *HOX2* probes were: a 3.0-kb *Sma*I genomic fragment of *HOX2E*; a 1.5-kb *Eco*RI–*Bgl*II fragment of *HOX2D*; a 1.3-kb cDNA clone (HHO.c1.95) from *HOX2C* (ref. 17); a 2.0-kb *Eco*RI fragment of *HOX2B*; a 1.2-kb cDNA clone (HHO.c10) from *HOX2A* (ref. 15); a 0.3-kb *Hind*III fragment of *HOX2F*; a 0.8-kb *Bgl*II–*Eco*RI fragment of *HOX2G*; a 2.8-kb *Eco*RI fragment of *HOX2H* and a 0.6-kb *Pvu*II fragment of *HOX2I*. The *HOX4B* probe was a 1.3-kb cDNA clone (HHO.c13)¹⁶.

FIG. 1 *a*, Expression of glycolipid surface antigen markers in the human EC cell line NT2/D1, before (C) and after induction with RA, at concentrations ranging 10^{-8} – 10^{-5} M. The EC stem cell markers SSEA-3, SSEA-4 and TRA1-60, and two antigens, ME311 and A2B5, which are expressed in EC-derived differentiated cells, were analysed using the corresponding monoclonal antibodies²². Neurofilament-positive neurons are observed among the A2B5⁺, ME311⁺ cell population. Mean $\pm$ s.e.m. of 3–6 different experiments are shown. *b*, Northern blot analysis of expression of *HOX2* genes and *c*, the *HOX4B* gene in NT2/D1, before (SC) and after 14 days of induction with 10^{-8} – 10^{-5} M RA. Transcript sizes are in kb. Open triangles, mobilities of 28S and 18S ribosomal RNAs used as size markers. Partial restriction maps of the respective clusters are shown at the bottom of each panel. *HOX* genes are identified according to the nomenclature for human homeobox genes²⁶. Related mouse designations indicated in parentheses. Black boxes, homeoboxes studied. E, *Eco*RI. Breaks shown in the *HOX2* and

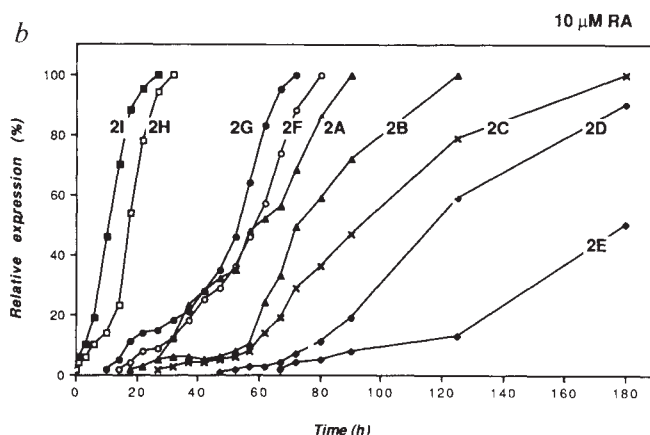
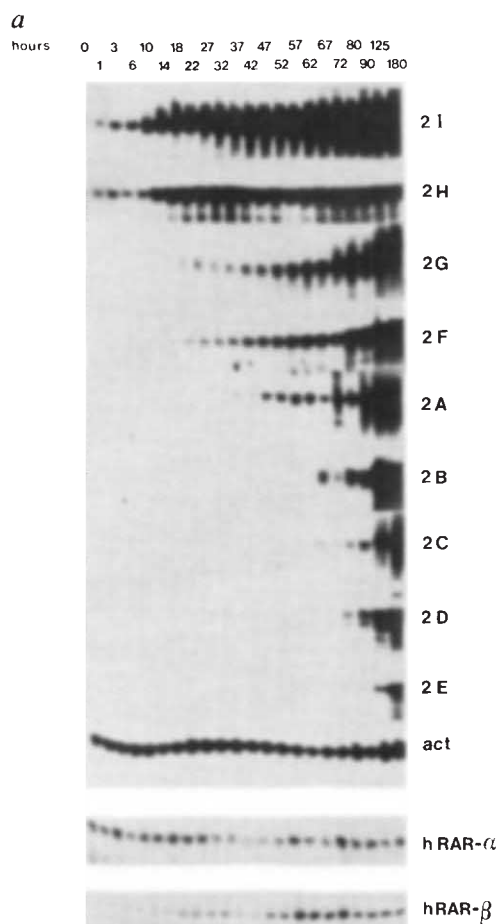


FIG. 2 **a**, RNase protection analysis of expression of the nine *HOX2* genes and human α - and β -retinoic acid receptor genes (hRAR- α , hRAR- β) in total RNA from NT2/D1 cells, uninduced (0) and induced with 10^{-5} M RA for 1–195 h. Expression of β -actin is an internal control. **b**, Values obtained for *HOX2* genes by densitometric scanning of autoradiographs, expressed in arbitrary units; 100 represents plateau RNA accumulation levels. These levels were comparable among different genes.

METHODS. RNA was obtained as described in the legend to Fig. 1. Partial 3' noncoding sequences from *HOX2* genes and from the human β -actin gene²⁹ were subcloned in pGEM vectors (Promega Biotec). A *Sma*–*Eco*R1 fragment from hRAR- α and an *Eco*R1–*Bam*HI fragment from hRAR- β cDNAs^{6,8} were subcloned in pBluescript vectors (Stratagene). Antisense strand RNA probes were synthesized with Sp6 or T7 polymerase and hybridized to 20 μ g RNA at appropriate temperatures. RNase digestion and electrophoresis on 7% urea-polyacrylamide gels were carried out as described previously³⁰.

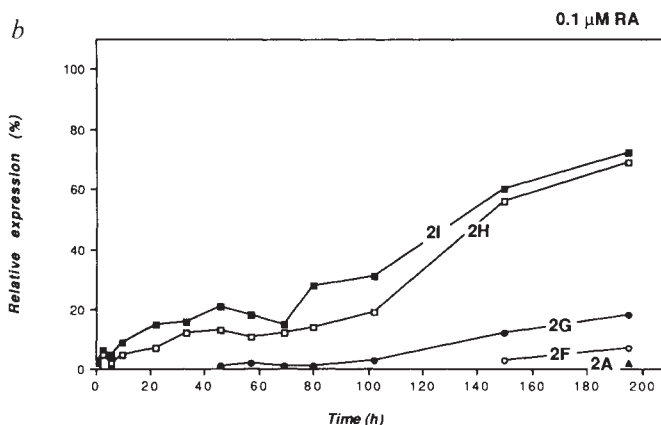
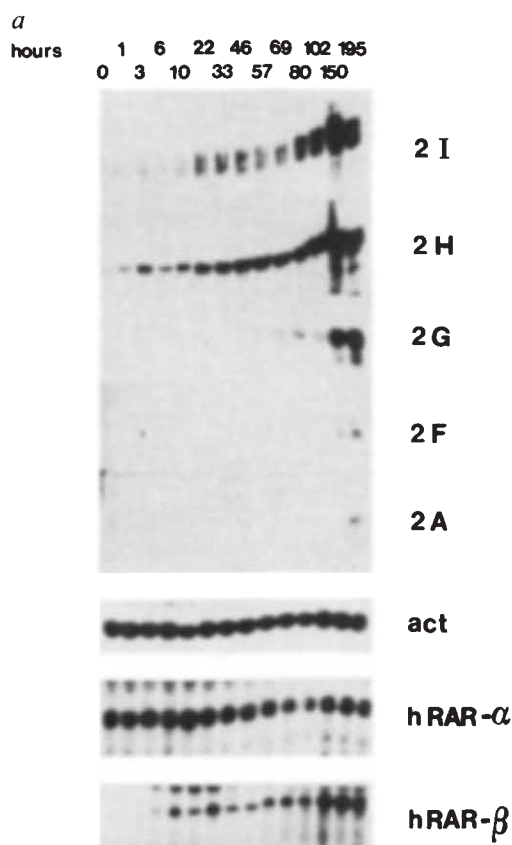


FIG. 3 **a**, RNase protection analysis of expression of *HOX2I*, *2H*, *2G*, *2F*, *2A*, hRAR- α , hRAR- β and β -actin genes in total RNA from NT2/D1 cells, before (0) and after induction with 10^{-7} M RA for 1–195 h. Expression of *HOX2B*, *2C*, *2D* and *2E* was undetectable up to 195 h (not shown). Values from autoradiograph densitometric scanning are shown in **b** (see legend to Fig. 2). Methods as for Fig. 2.

The time course of *HOX2* gene activation in cells induced by 10^{-7} M RA is shown in Fig. 3. Genes from *HOX2I* to *HOX2A* were activated in the same order observed in the presence of 10^{-5} M RA, although with a slower kinetics. Transcripts from genes upstream of *HOX2A* were undetectable throughout the period examined. It is important to note that NT2/D1 cells are morphologically and phenotypically homogeneous in the first week of culture, with no obvious difference between low and high RA levels²³. But we cannot rule out the possibility that 5'

genes, which are activated only at high RA levels, may be responding secondarily to differentiation events.

We also studied the expression of different RA receptors in NT2/D1 cells. Northern blot analysis showed transcripts for the α - and γ -receptors (hRAR- α and hRAR- γ) in both untreated and RA-treated EC cells, whereas hRAR- β transcripts were abundantly accumulated only in RA-treated cells, with the whole range of RA concentrations (data not shown). Appearance of hRAR- β transcripts slightly followed that of the early responders *HOX2I* and *HOX2H* (Figs 2a and 3a).

The level of RA needed to fully activate 5' *HOX2* genes in EC cells, 5×10^{-6} M appears to be higher than the levels reported in chick limb buds (10^{-8} – 10^{-7} M) (ref. 2) or *Xenopus* embryos (1.5×10^{-7} M) (ref. 5), although there is evidence that RA concentrations of the order of 10^{-6} M may be physiological in mammals²⁴. To estimate the range of effective RA concentrations in our experimental system, we cotransfected NT2 cells with an expression vector for human RAR- α (hRAR- α 0) and a reporter plasmid (TRE3)₃-tk-CAT containing a synthetic RA-responsive element⁹. Transfected cells were exposed to increasing RA concentrations and RA-induced activation was estimated by determining chloramphenicol acetyltransferase (CAT) activity expressed from the reporter gene (Fig. 4). Transactivation was observed starting from 10^{-10} M RA and steadily increased up to 10^{-5} M, with an estimated 50% maximum inducible CAT activity around values of 10^{-7} M. The range of effective concentrations in NT2 cells appears to be wider than previously observed in HeLa cells^{6,8,9}. The high levels of cytoplasmic RA-binding protein typical of human teratoma-derived cells²⁵ might at least in part account for this difference. A comparison between RA levels effective on cell lines *in vitro* and putative physiological values *in vivo* (as yet unknown in mammals) seems premature, and must await collection of more data *in vivo*.

We have shown that human *HOX2* genes are differentially activated in EC cells by RA in a concentration-dependent fashion and in a sequential order which is colinear with their 3' to 5' arrangement in the cluster. The same polarity is observed in the expression pattern of *HOX2* genes along the antero-posterior axis of the developing human and mouse CNS, where 3' genes are expressed more rostrally in myelencephalon and 5' genes more caudally in the spinal cord^{13,19–21}. The present data suggest

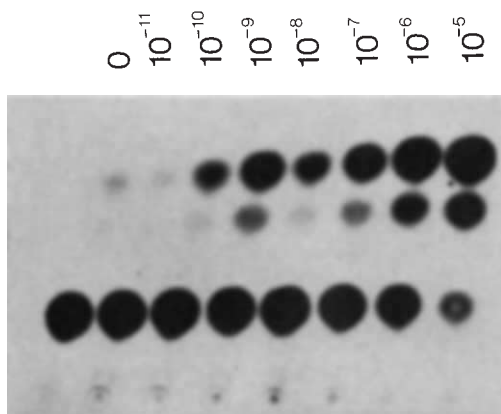


FIG. 4 RA-dependent transcriptional transactivation by hRAR- α in NT2/D1 cells. An expression vector containing the human RAR- α cDNA under the control of the simian virus 40 (SV40) early promoter (hRAR- α 0) was cotransfected with a reporter plasmid, (TRE3)₃-tk-CAT, containing a synthetic RA responsive element⁹. Cells were then exposed to increasing RA concentrations (10^{-11} – 10^{-5} M as shown) and assayed for transient CAT activity.

METHODS. NT2/D1 cells were cotransfected with hRAR- α 0 (5 μ g) and (TRE3)₃-tk-CAT (15 μ g), by a standard calcium phosphate technique³¹. After transfection (24 h) cells were fed with media containing different RA concentrations (10^{-11} – 10^{-5} M) or no RA (0). After 48 h of culture in RA-containing medium, cells were collected and cytosolic extracts prepared for CAT activity determination by TLC according to standard techniques³¹.

that temporal and/or spatial modulation of RA levels might, at least in part, underlie the observed spatially restricted *HOX2* gene expression patterns. This could in turn provide positional information required for further developmental decisions. It is interesting to note that *HOX2A*, which marks the border between genes requiring high or low levels of RA for their activation *in vitro*, is the homologue of mouse *Hox-2.1*, which marks the border between hindbrain and spinal cord *in vivo*¹⁹. □

Received 27 October 1989; accepted 7 June 1990.

1. Tickle, C. *et al.* *Nature* **296**, 564–565 (1982).
2. Thaller, C. & Eichele, G. *Nature* **327**, 625–628 (1987).
3. Thaller, C. & Eichele, G. *Development* **103**, 473–483 (1988).
4. Brookes, J.P. *Neuron* **2**, 1285–1294 (1989).
5. Durston, A. J. *et al.* *Nature* **340**, 140–144 (1989).
6. Petkovich, M., Brand, N.J., Krust, A. & Chambon, P. *Nature* **330**, 444–450 (1987).
7. Giguere, V., Ong, E. S., Sequi, P. & Evans, R. M. *Nature* **330**, 624–629 (1987).
8. Brand, N. J. *et al.* *Nature* **332**, 850–853 (1988).
9. Zelen, A., Krust, A., Petkovich, M., Kastner, P. & Chambon, P. *Nature* **339**, 714–717 (1989).
10. Mangelsdorf, D. J., Ong, E. S., Dyck, J. A. & Evans, R. M. *Nature* **345**, 224–229 (1990).
11. Gehring, W. J. *Science* **236**, 1245–1252 (1987).
12. Ingham, P. W. *Nature* **335**, 25–34 (1988).
13. Holland, P. W. H. & Hogan, B. L. M. *Genes Dev.* **2**, 773–782 (1988).
14. De Robertis, E. M., Oliver, G. & Wright, V. E. *Cell* **57**, 189–191 (1989).
15. Simeone, A. *et al.* *Nature* **320**, 763–765 (1986).
16. Mavilio, F. *et al.* *Nature* **324**, 664–668 (1986).
17. Simeone, A. *et al.* *Proc. natn. Acad. Sci. U.S.A.* **84**, 4914–4918 (1987).
18. Graham, A., Papalopulu, N. & Krumlauf, R. *Cell* **57**, 367–378 (1989).
19. Wilkinson, D. G., Bhatt, S., Cook, M., Boncinelli, E. & Krumlauf, R. *Nature* **341**, 405–409 (1989).
20. Giampaolo, A. *et al.* *Differentiation* **40**, 191–197 (1989).
21. Duboule, D. & Dollé, P. *EMBO J.* **8**, 1497–1505 (1989).
22. Mavilio, F., Simeone, A., Boncinelli, E. & Andrews, P. W. *Differentiation* **37**, 73–79 (1988).
23. Andrews, P. W. *Dev. Biol.* **103**, 285–293 (1984).
24. McCormick, A. M. & Napoli, J. L. *J. biol. Chem.* **257**, 1730–1735 (1982).
25. Matthaei, K. I., Andrews, P. W. & Bronson, D. L. *Expl. Cell Res.* **143**, 471–474 (1983).
26. Acampora, D. *et al.* *Nucleic Acids Res.* **17**, 10385–10402 (1989).
27. Chirgwin, J. M. *et al.* *Biochemistry* **18**, 5294–5300 (1979).
28. Thomas, P. S. *Proc. natn. Acad. Sci. U.S.A.* **77**, 5201–5205 (1980).
29. Ponte, P., Gunning, P., Blau, H. & Kedes, L. *Molec. cell Biol.* **3**, 1783–1791 (1983).
30. Melton, D. A. *et al.* *Nucleic Acids Res.* **12**, 7035–7056 (1984).
31. Sambrook, J., Fritsch, E. F. & Maniatis, T. *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor, New York, 1989).

ACKNOWLEDGEMENTS. We thank Dr M. Petkovich and Prof. P. Chambon for providing hRAR- α , hRAR- β and hRAR- γ cDNA sequences and hRAR- α 0 and (TRE3)₃-tk-CAT vectors. We also thank V. Nigro and M. Rambaldi for helpful suggestions, K. Wagner for technical help and N. Matti for typing the manuscript. This work has been supported by Associazione Italiana per la Ricerca sul Cancro (AIRC), the CNR Special Projects 'Biotechnology' and 'Human Genome Mapping', and by the NIH.

Reversible inactivation of K⁺ channels of *Vicia* stomatal guard cells following the photolysis of caged inositol 1,4,5-trisphosphate

Michael R. Blatt*†, Gerhard Thiel* & David R. Trentham‡

* Botany School, University of Cambridge, Downing Street, Cambridge CB2 3EA, UK

‡ National Institute for Medical Research, Mill Hill, London NW7 1AA, UK

RECENT investigations suggest that cytoplasmic D-myo-inositol 1,4,5-trisphosphate (InsP₃) functions as a second messenger in plants, as in animals, coupling environmental and other stimuli to intracellular Ca²⁺ release^{1,2}. Cytoplasmic levels of InsP₃ and the turnover of several probable precursors in plants are affected by physiological stimuli—including light, osmotic stress and the phytohormone indoleacetic acid^{3–5}—and InsP₃ activates Ca²⁺ channels⁶ and Ca²⁺ flux across plant vacuolar⁷ and microsomal membranes⁸. Complementary data also link changes in cytoplasmic free Ca²⁺ to several physiological responses, notably in guard cells which regulate gas exchange through the stomatal pores of higher plant leaves. Recent evidence indicates that guard cell K⁺ channels and, hence, K⁺ flux for stomatal movements⁹ may be

† Permanent address: Department of Biochemistry and Biological Sciences, University of London, Wye College, Wye (Ashford), Kent TN25 5AH, UK.

controlled by cytoplasmic Ca^{2+} (ref. 10). So far, however, direct evidence of a role for InsP_3 in signalling in plants has remained elusive. Here we report that InsP_3 released from an inactive, photolabile precursor, the P^5 -1-(2-nitrophenyl)ethyl ester of InsP_3 (caged InsP_3)¹¹ reversibly inactivates K^+ channels thought to mediate K^+ uptake by guard cells from *Vicia faba* L. while simultaneously activating an apparently time-independent, inward current to depolarize the membrane potential and promote K^+ efflux through a second class of K^+ channels^{12,13}. The data are consistent with a transient rise in cytoplasmic free Ca^{2+} (ref. 9) and demonstrate that intact guard cells are competent to use InsP_3 in signal cascades controlling ion flux through K^+ channels.

Caged InsP_3 was biologically inactive in the guard cells, as could be determined electrically. Bathed in 5 mM Ca^{2+} -HEPES buffer, pH 7.4, with 0.1 mM KCl, intact guard cells showed membrane potentials and resistances of -137 ± 4 mV and 12 ± 1 k Ω cm² ($n = 11$) when impaled with microelectrodes containing caged- InsP_3 (final concentration, 0.15 mM in 70 mM K^+ -acetate, pH 7.5). Equivalent results were obtained with microelectrodes filled with K^+ -acetate plus 0.15 mM caged ATP (the P^5 -1-(2-nitrophenyl)ethyl ester of ATP) or with K^+ -acetate alone, and in all cases the steady-state current-voltage (I - V) characteristics were dominated by an outward-rectifying K^+ channel current at positive-going voltages¹².

Raising the K^+ concentration of the bathing medium (K_0^+) to 30 mM depolarized the cells to -33 ± 2 mV ($n = 9$) and exposed a time-dependent, inward-going current which activated at voltages negative of -100 mV (Fig. 1a, b). The current was reversibly blocked by the K^+ channel antagonists tetraethylammonium chloride (30 mM) and 10 mM BaCl_2 (data not shown), and analyses of tail currents (Fig. 1c) revealed a Nernstian dependence on K_0^+ over the concentration range 3–30 mM (Fig. 1d),

indicating that this inward-rectifying current is carried by K^+ .

Light flashes on cells loaded with caged InsP_3 were followed by membrane depolarizations ($+15 \pm 2$ mV, $n = 16$) which subsequently recovered. Voltage trajectories varied somewhat from cell to cell (see Figs 2–4); in all but three trials, however, depolarizations were complete within 2 min (mean $\pm$ s.e., 0.9 ± 0.3 min). Figure 2 shows that, underlying this voltage response, the effect was to suppress the time- and voltage-dependent inward (K^+ channel) current while increasing the instantaneous (leak) current. By contrast, the outward-rectifying K^+ current showed no measurable response in seven cells; in four cells it was reduced in scalar fashion but recovered within 2–3 min (compare Fig. 3).

Figure 3 indicates a voltage-dependence to the response of the inward-rectifying K^+ current. Flash-evoked inactivation of the current was marked both by a decline in the current magnitude, and by a shift of around -50 mV in the voltage at which 50% activation occurred (Fig. 3b). These same characteristics are evident in Fig. 4 which shows that a graded response of the inward-rectifier could be achieved with increasing InsP_3 release (flash intensity). Figure 4 also shows that raising the cytoplasmic Ca^{2+} -buffering capacity, by including 20 mM EGTA in the microelectrode, was antagonistic to the InsP_3 -evoked response (compare Fig. 4c and d).

Recovery of the K^+ channel and leak currents—and of membrane voltage—was correlated with flash exposure. Focusing flashes on impaled cells using the camera port of the microscope resulted in current and voltage responses which recovered fully over 5–8 min (Fig. 3); higher beam intensities could be achieved by mounting the flashlamp beside the microscope and directing the beam on the chamber, but with the disadvantage that the caged compound in the microelectrode was exposed also;

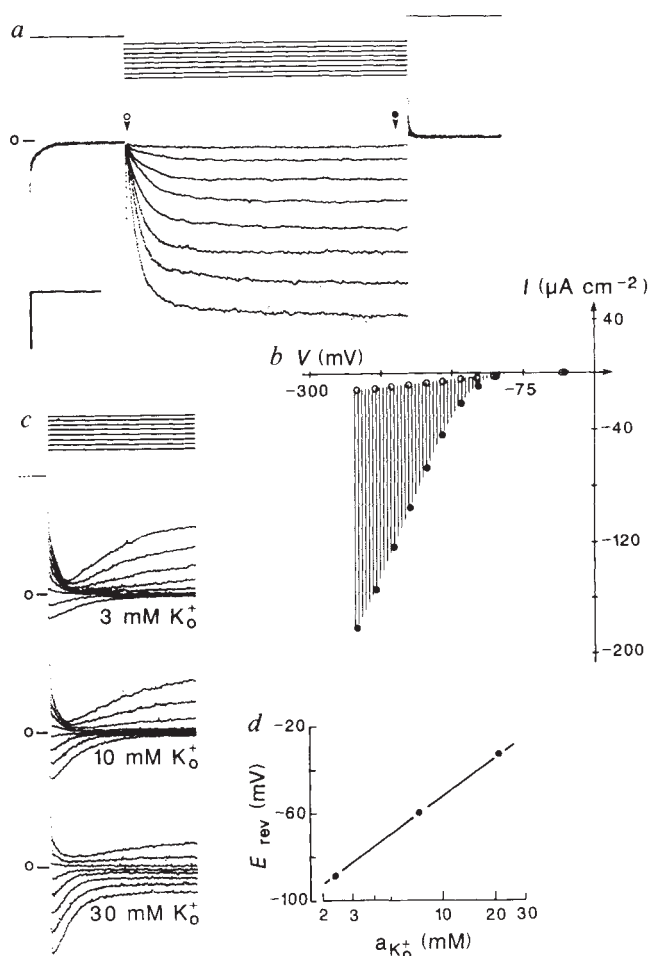


FIG. 1 Inward-rectifying K^+ channel current (I) of intact *Vicia* guard cells. Data from one cell: surface area, 1.6×10^{-5} cm²; volume, 3.4 pL; stomatal aperture, 6 μm . a, Kinetics of voltage-dependent activation in 30 mM K_0^+ . Eight cycles of three voltage-clamp steps (above): conditioning voltage, -100 mV; test voltages, -120 to -250 mV; tailing voltage, -30 mV. Current zeros indicated to the left. Scale bar: vertical, 60 $\mu\text{A cm}^{-2}$ or 200 mV; horizontal, 500 ms. b, Current-voltage characteristics of the instantaneous ($\circ$, leak) and steady-state ($\bullet$, K^+ channel plus leak) currents from traces in a (sampling times cross-referenced by symbol; steady-state current at -100 mV included). Shaded difference is the K^+ current. c, Current 'tails' recorded after 1 s conditioning steps to -250 mV. Above, clamp voltage steps (8: 0 to -120 mV); below, tail currents recorded in 3, 10 and 30 mM K_0^+ . Current zeros indicated to the left. Scale bar: vertical, 40 $\mu\text{A cm}^{-2}$ or 200 mV; horizontal, 200 ms. The late rising currents at more positive voltages reflect activation of the outward-rectifying K^+ current¹². d, Dependence of tail current reversal potential (E_{rev}) on extracellular K^+ activity ($a_{\text{K}_0^+}$). E_{rev} taken from current relaxations in c and K_0^+ corrected for activities in free solution (see ref. 12). Slope of the line, $+59$ mV/ $a_{\text{K}_0^+}$ decade.

METHODS. *Vicia faba* L. was grown and epidermal peels containing the guard cells were mounted as described previously^{13,20}. Peels were bathed in fast-flowing solutions containing KCl and 5 mM Ca^{2+} -HEPES buffer, pH 7.4, or 5 mM Ca^{2+} -MES, pH 6.1^{12,14}. Electrical recordings were carried out using double-barrelled microelectrodes²⁰ filled with 200 mM K^+ -acetate^{12,14}. Surface areas were calculated from the orthogonal dimensions of impaled cells^{12,14}. Cells were loaded with the caged compounds, inositol phosphates and Ca^{2+} buffers by diffusion from the microelectrode (voltage-recording barrel) and the K^+ -acetate filling solution was doped accordingly. Estimates of loading times based on calculated and measured electrolyte leakage rates from microelectrodes^{14,21} indicated that the compounds should equilibrate with the cell cytoplasm within 15–20 min; trials with the caged compounds generally were begun 15 min, and in some experiments with caged ATP, 25–30 min after impalement. Caged InsP_3 and caged ATP were photolysed by 1-ms flashes from a xenon flashlamp (Chadwick Helmholtz, El Monte, USA). Photochemical release was calibrated in trials with 5 μM caged ATP placed in the experimental chamber and the ATP released was measured by HPLC²² on a Parasil 10SAX column with a NaH_2PO_4 gradient. A single, full-intensity flash (55 mJ, effective wavelengths 300–350 nm recorded through a UG11 filter) released 9% of the caged compound and this value was used also in estimating InsP_3 release¹¹. Because of uncertainties about diffusional loading from the microelectrode²¹, however, the estimates should be treated as maxima only.

recoveries in these instances were slow (10–20 min) or incomplete (Figs 2, 4), probably a consequence of InsP_3 release in the microelectrodes and its subsequent diffusion into the cells.

The flash-dependent effects on K^+ channel and leak currents were specifically associated with the release of InsP_3 . Both the current and voltage responses were reproducible and could be repeated in any one guard cell impaled with a microelectrode containing caged InsP_3 ; comparable electrical characteristics were obtained with microelectrodes containing $10\text{ }\mu\text{M}$ native InsP_3 ($n=4$) but not with *D-myo*-inositol 1,4-bisphosphate ($n=3$). Furthermore, light flashes failed to evoke any response in cells ($n=13$) impaled with microelectrodes containing 0.15 mM caged-ATP. The latter observation provides strong evidence also that neither the light flash itself, nor the by-products from photolysis of caged InsP_3 (2-nitrosoacetophenone and H^+ (ref. 11)) had any short-term or lasting effect on the cells.

We note that a physiological effect of InsP_3 on transport across the plasma membrane is easily explained only if caged InsP_3 were loaded in the cytoplasm, that is, if, in each case, the microelectrode tip was located in the cytoplasmic compartment. This interpretation accords with the results of dye injections¹⁰ and Cl^- leakage from microelectrodes¹⁴ into *Vicia* guard cells, and with reports of cytoplasmic pH and pCa values recorded

with ion-sensitive microelectrodes^{15,16}. We conclude, as have others¹⁰, that physiologically viable impalements of higher plant cells—in our hands, those that yield stable electrical characteristics over many tens of minutes—leave the microelectrode tip in the cytoplasm.

The primary response to InsP_3 release in these experiments was most probably a rise in cytoplasmic free Ca^{2+} . Estimates of the jump in InsP_3 generated by a flash (see legend to Fig. 1) lie in the range $0.5\text{--}10\text{ }\mu\text{M}$, concentrations effective in releasing Ca^{2+} from plant vacuoles¹⁷. Furthermore, the effects of InsP_3 release were countered by raising the cytoplasmic Ca^{2+} -buffering capacity, although channel activity in the control appeared unaffected (Fig. 4d). There is, also, a remarkable similarity between our data and those reported for *Vicia* guard cell protoplasts loaded with Ca^{2+} (ref. 9): in each case, the effect was to reduce inward-rectifying K^+ currents and alter channel gating,

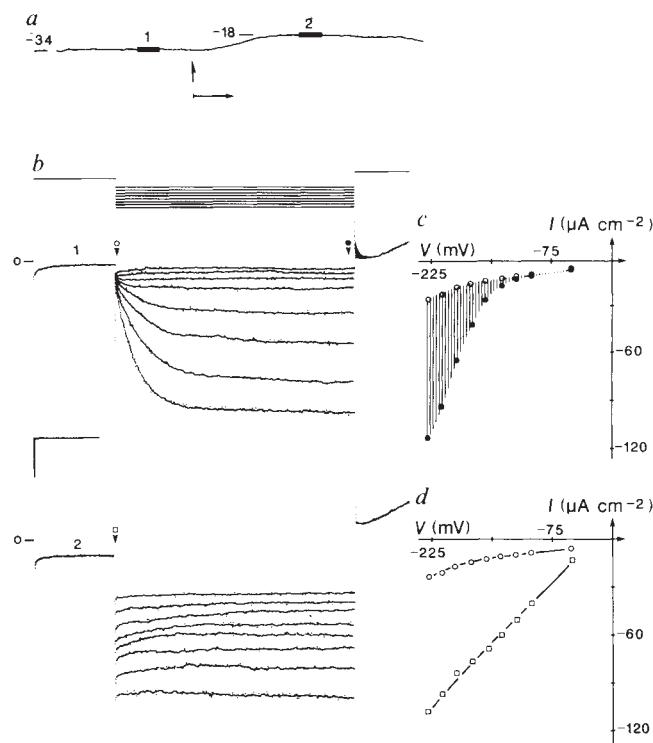


FIG. 2 K^+ channel and leak current response to photolytic release of InsP_3 from caged InsP_3 . Data from one *Vicia* guard cell loaded with 0.15 mM caged InsP_3 and bathed in 5 mM Ca^{2+} -MES, pH 6.1, with 30 mM KCl . Cell parameters: surface area, $1.6 \times 10^{-5}\text{ cm}^2$; volume, 4.1 pL ; stomatal aperture, $7\text{ }\mu\text{m}$. *a*, Free-running membrane potential (values in mV) before and after flash ($\uparrow$, 25 mJ ; estimated InsP_3 release, $5\text{ }\mu\text{M}$). Voltage clamp cycles (masked from trace) were run during the barred periods. Timescale, 1 min. *b*, Current trajectories recorded during the periods (1, before; 2, after flash) indicated in *a*. Current zeros indicated to left. Clamp cycle (above): conditioning voltage, -50 mV ; test voltages (δ), -100 to -250 mV ; tailing voltage, 0 mV . Scale bar: vertical, $30\text{ }\mu\text{A cm}^{-2}$ or 300 mV ; horizontal, 300 ms . Instantaneous ($\circ, \square$) and steady-state ($\bullet$) currents taken at times indicated (cross-referenced to *c, d* by symbol). Note the absence of the time-dependent (K^+) current after the flash. *c*, Current-voltage relations for the instantaneous ($\circ$; leak) and steady-state ($\bullet$; K^+ channel plus leak) currents before the flash (steady-state current at -50 mV included). Shading indicates the time-dependent (K^+ channel) component. *d*, Instantaneous current-voltage relations before ($\circ$) and after ($\square$) flash, showing around fivefold rise in leak conductance.

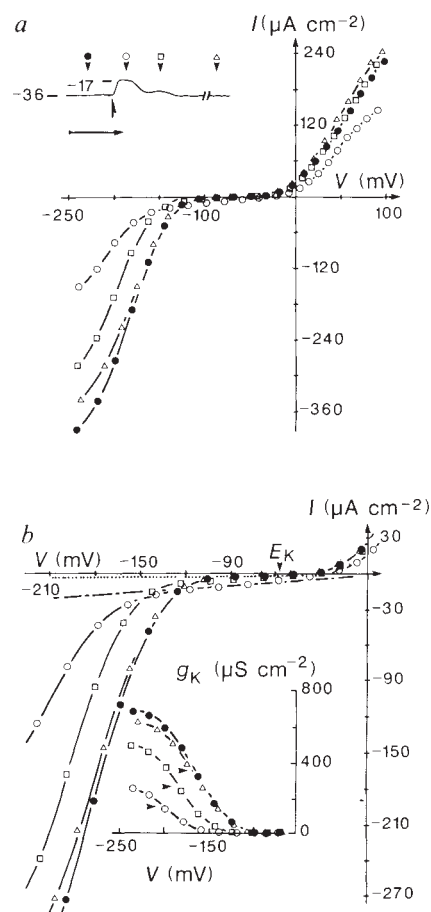


FIG. 3 Time course for K^+ channel and leak current response and recovery following photolytic release of InsP_3 . Data from one *Vicia* guard cell loaded with 0.15 mM caged InsP_3 and bathed in 5 mM Ca^{2+} -MES, pH 6.1, with 10 mM KCl . Cell parameters: surface area, $1.6 \times 10^{-5}\text{ cm}^2$; volume, 3.6 pL ; aperture $8\text{ }\mu\text{m}$. *a*, Current-voltage curves determined before ($\bullet$) and after ($\circ, \square, \Delta$) flash through microscope camera port ($\uparrow$, 5 mJ ; estimated InsP_3 release, $1\text{ }\mu\text{M}$). Current-voltage scans run in a bipolar staircase protocol^{13,20} with 400-ms pulses. Scan duration, 16 s . Inset: Voltage trace (values in mV) with times of I - V scans (masked from trace) cross-referenced by symbol (final scan, 7.6 min after flash). Timescale, 2 min . *b*, Detail of data in *a* showing leak currents. E_K , K^+ equilibrium potential. The leak in the control ($\cdots$, $9\text{ }\mu\text{S cm}^{-2}$) was identified from instantaneous currents (not shown) as in Fig. 1. This approach was not possible within the short response time course of cells irradiated through the microscope camera port; instead, the voltage range over which the leak and steady-state I - V scans overlapped in the control was noted and, following flashes, steady-state currents recorded over this same voltage range were used to estimate leak currents ($---$, $34\text{ }\mu\text{S cm}^{-2}$) by linear extrapolation. Inset: Steady-state K^+ conductance (g_K) determined as $I_K/(V-E_K)$ after subtracting leak currents. Arrows mark 50% activation.

while promoting a time-independent, inward current. (Note that in the present circumstances, with $E_{Cl} \ll 0$ mV, a Cl^- flux^{9,18} could not have accounted for this additional $InsP_3$ -activated current.) Recovery following the light flashes is readily understood, by analogy with animal cell models, in context of subsequent $InsP_3$ metabolism and Ca^{2+} sequestration¹.

Finally, the results of the present study allow us to piece together independent reports on the action of abscisic acid (ABA), a phytohormone that promotes stomatal closure by inducing guard cells to lose osmotica, notably K^+ salts¹⁹. Exposing *Vicia* guard cells to ABA results in the rapid decline of an

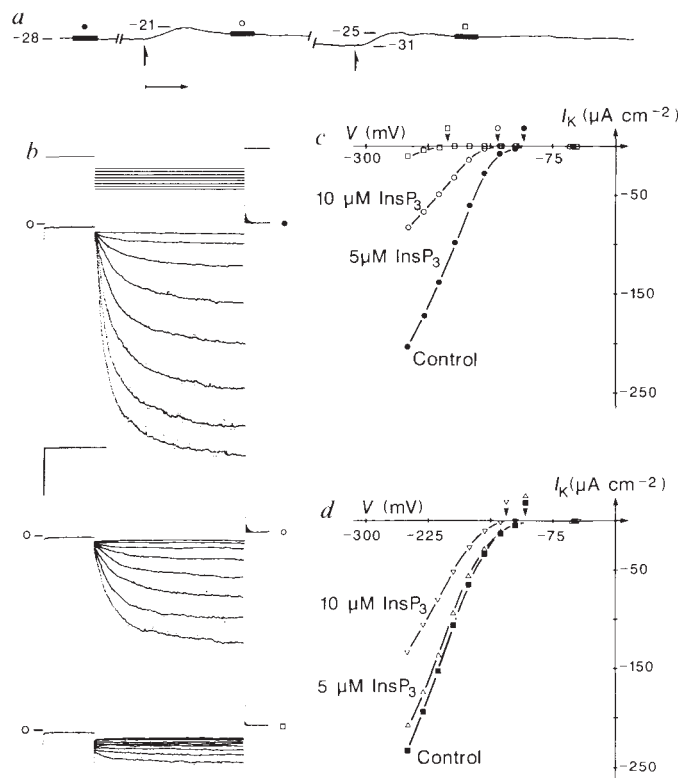


FIG. 4 Cumulative response of the K^+ current to photolytic release of $InsP_3$ and its block by buffering the cytoplasmic free Ca^{2+} concentration. Data from *Vicia* guard cells bathed in 5 mM Ca^{2+} -MES, pH 6.1, with 30 mM KCl, one cell (a–c) loaded with 0.15 mM caged- $InsP_3$, the second cell (d) loaded with 0.15 mM caged- $InsP_3$ plus 20 mM K_2 EGTA, pH 7.5 ($CaCl_2$ added to $pCa \approx 80$ nM²³). Cell parameters (both cells): surface area, 1.5×10^{-5} cm²; volume, 3.4 pL; stomatal aperture 6 μ m. a, Free-running membrane potential (values in mV) before and after flashes (|; first flash 25 mJ ($InsP_3$ released, 5 μ M), 18 min after impalement; second flash 55 mJ ($InsP_3$ released, 10 μ M), 31 min after impalement; total duration of impalement, 51 min). Voltage clamp cycles (masked from trace) were run during the barred periods (cross-referenced to b, c by symbol). Timescale, 1 min. b, Current trajectories recorded during the periods indicated in a. Current zeros indicated to left. Clamp cycle (above): conditioning voltage, –50 mV; test voltages (8), –120 to –250 mV; tailing voltage, 0 mV. Scale bar: vertical, 50 μ A cm^{–2} or 300 mV; horizontal, 500 ms. No effect of $InsP_3$ on the kinetics for activation by voltage was observed. Between flashes the K^+ current varied by less than 20%. c, Steady-state K^+ currents (including current at –50 mV) from b after subtracting the instantaneous currents (see Fig. 1). Note the cumulative shift of around –100 mV in activation potential (carats, cross-referenced by symbol) recorded for the K^+ current. d, Steady-state K^+ currents from a second cell as in c, but after loading with 20 mM EGTA to buffer $InsP_3$ -induced changes in cytoplasmic free Ca^{2+} concentration. Symbols: ■, control; △, following first flash (25 mJ; $InsP_3$ released, 5 μ M), 17 min after impalement; ▽, following second flash (55 mJ; $InsP_3$ released, 10 μ M), 32 min after impalement; total duration of impalement, 39 min. No effect of the buffer on current activation kinetics nor, before $InsP_3$ release, on membrane potential or steady-state current (not shown) was observed. Comparable results were obtained from two other cells impaled with EGTA-containing microelectrodes, and also when the Ca^{2+} buffer BAPTA²⁴ was substituted for EGTA.

inward-rectifying current¹³, which we now show to be carried by K^+ (Fig. 1), and a coincident rise in an unidentified leak current. Again, the behaviour parallels guard cell membrane response to Ca^{2+} . The effects may be seen to prevent K^+ uptake through the inward-rectifying channels while activating the leak to depolarize the membrane and promote K^+ loss through outward-rectifying K^+ channels^{12,13}. ABA also raises the free Ca^{2+} concentration of the guard cell cytoplasm¹⁰. Taken together, these observations offer strong support for the action of ABA on the inward-rectifying K^+ channel and leak currents being effected by cytoplasmic Ca^{2+} .

Our findings clearly demonstrate that guard cells are competent to use $InsP_3$ in controlling K^+ channel activity. The data now also raise the possibility that $InsP_3$ -mediated Ca^{2+} release forms another link in the ABA transduction chain. An immediate issue is whether the phytohormone promotes $InsP_3$ turnover. There is, further, a need to address the possible concurrent functioning of other signalling pathways in the guard cells; indeed, ABA also activates outward-rectifying K^+ channels in the guard cells¹³, and neither $InsP_3$ nor cytoplasmic Ca^{2+} (ref. 9) seems to have a significant effect on these currents. □

Received 26 March; accepted 15 June 1990.

- Berridge, M. J. & Irvine, R. F. *Nature* **341**, 197–202 (1989).
- Thiel, G., MacRobbie, E. A. C. & Hanke, D. *EMBO J.* **9**, 1737–1741 (1990).
- Morse, M. J., Crain, R. C. & Satter, R. L. *Proc. natn. Acad. Sci. U.S.A.* **84**, 7075–7078 (1987).
- Ettlinger, C. & Lehle, L. *Nature* **331**, 176–178 (1988).
- Srivastava, A., Pines, M. & Jacoby, B. *Physiol. Pl.* **77**, 320–325 (1989).
- Alexandre, J., Lassalles, J. P. & Kado, R. T. *Nature* **343**, 567–570 (1990).
- Ranjana, R., Carrasco, A. & Boudet, A. M. *FEBS Lett.* **230**, 137–141 (1988).
- Drabak, B. K. & Ferguson, I. B. *Biochem. Biophys. Res. Commun.* **130**, 1241–1246 (1985).
- Schroeder, J. I. & Hagiwara, S. *Nature* **338**, 427–430 (1989).
- McAinsh, M. R., Brownless, C. & Hetherington, A. M. *Nature* **343**, 186–188 (1990).
- Walker, J. W., Feeney, J. & Trentham, D. R. *Biochemistry* **28**, 3272–3280 (1989).
- Blatt, M. R. *J. Membrane Biol.* **102**, 235–246 (1988).
- Blatt, M. R. *Planta* **180**, 445–455 (1990).
- Blatt, M. R. *Planta* **170**, 272–287 (1987).
- Miller, A. J. & Sanders, D. *Nature* **326**, 397–400 (1987).
- Felle, H. *Planta* **174**, 495–499 (1988).
- Brosnan, J. M. & Sanders, D. *FEBS Lett.* **260**, 70–72 (1990).
- Keller, B. U., Hedrich, R. & Raschke, K. *Nature* **341**, 450–453 (1989).
- MacRobbie, E. A. C. in *Stomatal Function* (eds Zeiger, E., Farquhar, G. D. & Cowan, I. R.) 125–162 (Stanford University Press, Stanford, California, 1987).
- Blatt, M. R. *J. Membrane Biol.* **98**, 257–274 (1987).
- Blatt, M. R. & Slayman, C. L. *J. Membrane Biol.* **72**, 223–234 (1983).
- Walker, J., Reid, G. & Trentham, D. R. *Meth. Enzym.* **172**, 288–301 (1989).
- Fabiato, A. & Fabiato, F. *J. Physiol., Paris* **75**, 463–505 (1979).
- Pethie, R. et al. *Cell Calcium* **10**, 491–497 (1989).

ACKNOWLEDGEMENTS. We thank Dr C. A. Brearley for help with the HPLC assays, and Dr D. Hanke and Professor E. A. C. MacRobbie for their support. The work was supported by the Gatsby Charitable Foundation. G.T. was supported by the British Science and Engineering Research Council.

Elevation of cytoplasmic calcium by caged calcium or caged inositol trisphosphate initiates stomatal closure

S. Gilroy, N. D. Read & A. J. Trewavas

Department of Botany, University of Edinburgh, The King's Buildings, Mayfield Road, Edinburgh EH9 3JH, UK

STOMATAL pores at the leaf surface regulate water loss during transpiration and CO_2 uptake for photosynthesis through changes in the turgor of the surrounding guard cells. Such regulation occurs in response to a wide range of environmental stimuli (such as light, CO_2 level and abscisic acid). These are thought to be transduced through changes in cytosolic calcium levels^{1–3}, and several treatments causing stomatal closure are accompanied by increases in cytosolic Ca^{2+} (ref. 4), probably from mobilization of intracellular Ca^{2+} stores. We report here the use of the fluorescent Ca^{2+} indicator Fluo-3 (refs 5, 6) to follow changes in stomatal aperture and cytoplasmic Ca^{2+} concentration as Ca^{2+} or inositol 1,4,5-

trisphosphate (InsP_3) are released into the cell from their caged (photoactivatable) forms⁶⁻⁹. Increasing Ca^{2+} concentration to greater than ~ 600 nM induced stomatal closure. Similarly, release of InsP_3 initiated influx of Ca^{2+} into the cytosol which was followed by stomatal closure. These results suggest that Ca^{2+} and InsP_3 may act as second messengers leading to guard cell closure.

Fluo-3 is a fluorescent Ca^{2+} indicator which has been successfully used to monitor changes in cytoplasmic Ca^{2+} levels ($[\text{Ca}^{2+}]_i$) in animal cells⁷. Increases in $[\text{Ca}^{2+}]_i$ lead to increases in Fluo-3 fluorescence intensity. This allows approximate calibration of $[\text{Ca}^{2+}]_i$ (refs 5, 6). The dye does not show a spectral shift on binding Ca^{2+} and so cannot be used for ratio analysis of Ca^{2+} levels¹⁰. Fluo-3 can, however, be used to monitor continuously $[\text{Ca}^{2+}]_i$ in cells loaded with caged probes as its excitation wavelength (490 nm) does not interfere with the cages. The caged probes are photolysed by the ultraviolet light used to excite the fluorescence of ratio Ca^{2+} probes such as Indo-1 or Fura-2 (ref. 10). Figure 1 shows typical traces of guard cells

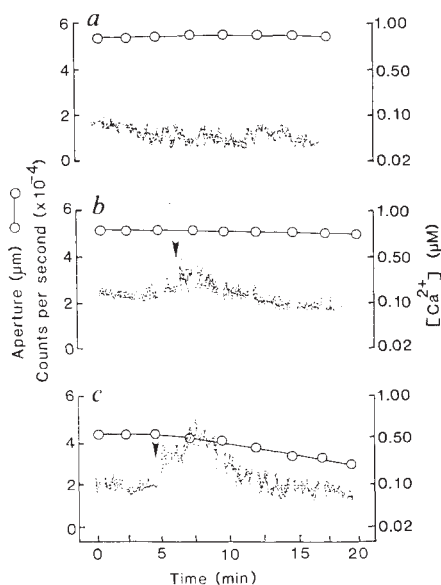


FIG. 1 Changes in $[\text{Ca}^{2+}]_i$ in guard cells of *Commelina communis* induced upon 0 s (a), 5 s (b) and 10 s (c) of ultraviolet photolysis (arrow) of Nitr-5 (caged Ca^{2+}).

METHODS. Abaxial (lower) epidermis of *C. communis* was peeled and fixed to a coverslip with small amounts of silicone grease around its periphery. The preparation was mounted in a microscope chamber and perfused with 50 mM KCl, 1 mM MES, 20 μM CaCl_2 , pH 6.15, at 25 °C. Guard cells were impaled with micropipettes (tip diameters <300 nm) filled with 100 μM Fluo-3 (Molecular Probes Inc.) and 100 μM tetrasodium Nitr-5 (Calbiochem), 100 μM CaCl_2 and iontophoretically microinjected at less than 0.1 nA for 5–10 min. The micropipette was then removed with no obvious loss of guard cell turgor and cells incubated for 15 min to allow recovery from the microinjection procedure. Fluo-3 fluorescence (excitation 490 nm, 10 nm half bandwidth + ND2 quartz neutral density filter; emission 530 nm, half bandwidth 10 nm; dichroic mirror 505 nm; filters supplied by Ealing Electro-Optics, Watford) was measured using a Nikon Diaphot inverted microscope, Nikon CF Fluor DL 40×, 0.85 numerical aperture, 0.37 mm working distance, dry objective and a simultaneous dual emission wavelength photometry system (Newcastle Photonic Systems, Newcastle-upon-Tyne). Caged probe was photolysed using ultraviolet light from the 75 W xenon epifluorescence illuminator passed through a 350 nm (10 nm half bandwidth) interference filter (Ealing Electro-Optics). Photolysis of the caged probes did not affect guard cell autofluorescence. A 10-s ultraviolet pulse photolysed $\sim 20\%$ of the caged Ca^{2+} *in vitro*. Fluo-3 fluorescence was calibrated for $[\text{Ca}^{2+}]_i$ at the end of each experiment using 1 mM Mn^{2+} and 10 μM Br-A23187 (ref. 7). However, calibration of Fluo-3 fluorescence intensity against absolute $[\text{Ca}^{2+}]_i$ (refs 6, 7) is extremely difficult and the values given should be taken as approximations only. The aperture of the guard cell in which the $[\text{Ca}^{2+}]_i$ was monitored was measured to 0.5 μm accuracy from micrographs taken at 2.5-min intervals throughout the experiment. Traces are representative of measurements on 10, 15 and 22 cells (Fig. 1a–c) respectively.

simultaneously microinjected with Fluo-3 and Nitr-5 (caged Ca^{2+}) (refs 6–8). The loaded guard cells were exposed to pulses of ultraviolet light of varying duration to photolyse different amounts of the caged Ca^{2+} and hence increase $[\text{Ca}^{2+}]_i$ to different levels. The loading of these cells with the caged probe did not lead to changes in stomatal aperture or $[\text{Ca}^{2+}]_i$ (Fig. 1a). When, however, $[\text{Ca}^{2+}]_i$ was elevated to greater than ~ 600 nM by ultraviolet photolysis of the caged Ca^{2+} , stomatal closure subsequently occurred ($n=27$) (Fig. 1b, c). Thus a threshold of Ca^{2+} activation for the ion fluxes leading to stomatal closure is indicated. In guard cells injected with only Fluo-3, or with Fluo-3 and Nitr-5 which had been completely photolysed with ultraviolet light, application of 10- to 15-second ultraviolet pulses did not affect stomatal aperture or $[\text{Ca}^{2+}]_i$. This suggests that the observed response of stomatal closure is unlikely to be due to side effects of the photolysis procedure.

Ca^{2+} -activated ion efflux channels have been identified in guard cell plasma and vacuolar membranes^{11–13}. These require 10^{-6} M Ca^{2+} for activation, and may represent part of the mechanism whereby elevation of $[\text{Ca}^{2+}]_i$ leads to stomatal closure. The sustained nature of the $[\text{Ca}^{2+}]_i$ increase induced by caged Ca^{2+} photolysis (Fig. 1) more closely resembles that reported for invertebrate giant axons than the quick changes seen in mammalian cells⁶. This may reflect the release of Ca^{2+} from intracellular stores induced by the elevated $[\text{Ca}^{2+}]_i$, paralleling the Ca^{2+} -induced Ca^{2+} release seen in animal cells¹⁴.

Stimulus-evoked Ca^{2+} release from internal stores may be mediated through the release of InsP_3 (ref. 14) and stimulation of $^{45}\text{Ca}^{2+}$ efflux from plant microsomal and vacuolar prepar-

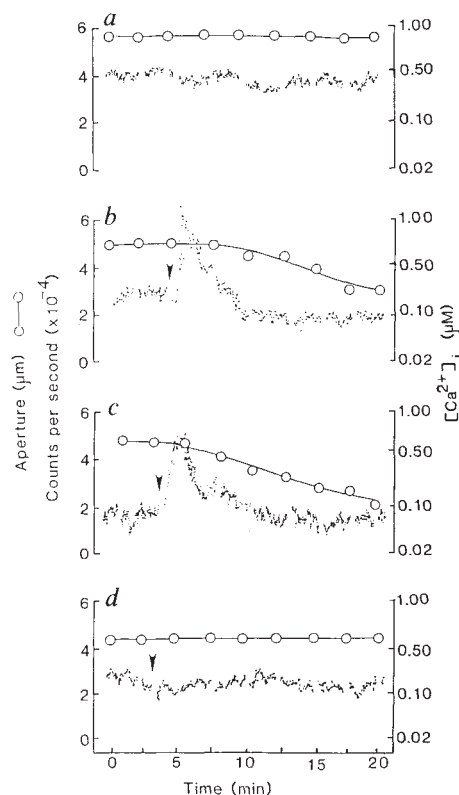
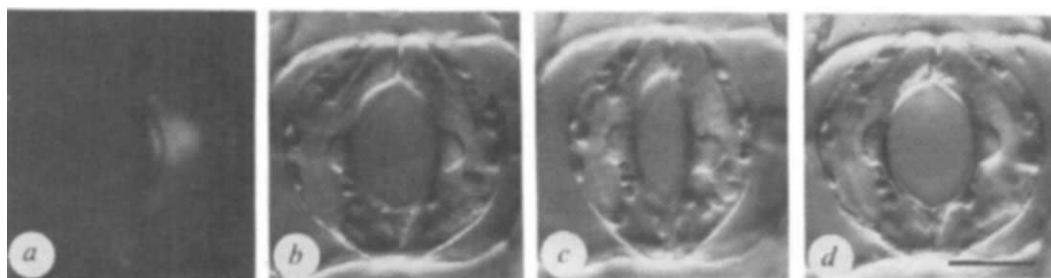


FIG. 2 Changes in $[\text{Ca}^{2+}]_i$ in guard cells of *C. communis* upon release of InsP_3 from its caged form. Cells were microinjected with caged InsP_3 and Fluo-3 as described in Fig. 1 except that the injection pipette contained 100 μM Fluo-3 and 100 μM trisodium *myo*-inositol-1,4,5-trisphosphate, $\text{P}^{4(5)}$ -1-(2-nitrophenyl)ethyl ester (caged InsP_3) (a–c) or caged ATP (d). Loaded cells were subjected to 0 s (a) or 10 s (b–d) of ultraviolet. The cell in b was preincubated for 15 min in 1 mM La^{3+} . Traces are representative of measurements on 10, 18, 8 and 12 cells (a–d) respectively. Cytoplasmic Fluo-3 concentration was estimated at 10–50 μM from comparison of the Ca^{2+} -saturated Fluo-3 signal in guard cells treated with 10 μM Br-A23187 to that of 6 pl droplets (guard cell sized) of a range of [Fluo-3] in silicone oil.

FIG. 3 Guard cell of *C. communis* microinjected with Fluo-3 and caged Ca^{2+} . *a*, Fluorescence micrograph of injected guard cell (excitation 475–505 nm, emission 520–550 nm). *b–d*, Differential interference contrast light micrographs of guard cell pair (right-hand cell only injected with Fluo-3 and caged Ca^{2+}) taken at 0 min (*b*) and 45 min (*c*) after 10 s photolysis of the caged Ca^{2+} and *d*, 120 min after the application of $10\ \mu\text{M}$ fusicoccin to the guard cell in *c*. Scale bar, $10\ \mu\text{m}$.



ations by InsP_3 has been reported^{15,16}. InsP_3 -activated Ca^{2+} channels have also been identified in the vacuole of red beet by patch clamping¹⁷. We have recently found, using fluorescence ratio imaging of Indo-1, that stimuli causing stomatal closure lead to localized increases in $[\text{Ca}^{2+}]_i$, probably due to Ca^{2+} release to the cytosol from organelles (S.G., M. D. Fricker, N.D.R. and A.J.T., unpublished results). This suggested that a second messenger such as InsP_3 may be involved in the transduction of these closure signals. InsP_3 was therefore released to the guard cell cytoplasm from its caged form (inositol-1,4,5-trisphosphate-nitrophenylethyl ester, NPE-caged- InsP_3) and the effects on stomatal aperture and $[\text{Ca}^{2+}]_i$ dynamics investigated.

Loading cells with caged InsP_3 did not induce changes in stomatal aperture or $[\text{Ca}^{2+}]_i$ (Fig. 2*a*). Release of free InsP_3 from the cage led to stomatal closure (Fig. 2*b*) however. It also resulted in a rise in $[\text{Ca}^{2+}]_i$ over 10 s which was sustained for 5–10 min, and which preceded stomatal closure (Fig. 2*b*). Co-injecting EGTA with the caged InsP_3 (at $100\ \mu\text{M}$ in the micro-pipette) abolished these effects ($n=8$) (data not shown). An increase in $[\text{Ca}^{2+}]_i$ that preceded stomatal closure was, however, still observed when the caged InsP_3 was photolysed in guard cells pretreated for 15 min with $1\ \text{mM}$ La^{3+} (Fig. 2*c*). La^{3+} inhibits Ca^{2+} influx at the guard cell plasma membrane³, suggesting that InsP_3 was mobilizing Ca^{2+} from internal stores. Equivalent experiments in which NPE-caged-ATP replaced the NPE-caged- InsP_3 did not result in stomatal closure or increases in $[\text{Ca}^{2+}]_i$ ($n=12$) (Fig. 2*d*), suggesting that the InsP_3 , rather than the photolysis procedure, was responsible for the observed effects. Similarly, directly injecting inositol 1,4,5-trisphosphate into guard cells seems to lead to stomatal closure, whereas inositol 2,4,5-trisphosphate does not (not shown). Thus InsP_3 does seem to be acting as a messenger inducing Ca^{2+} influx into the cytosol and initiating stomatal closure. The precise intracellular site of this Ca^{2+} release is being investigated. Our previous unpublished observations using fluorescence ratio imaging of Indo-1 suggest the endoplasmic reticulum and vacuole as prime candidates. Nuclear Ca^{2+} seems extremely stable despite large changes in cytosolic levels. Thus despite its large Fluo-3 signal (Fig. 3), the variations in $[\text{Ca}^{2+}]_i$ we have observed seem less likely to have arisen in the nucleus than the cytosol.

The photolysis products of caged compounds¹⁸ have been reported to be cytotoxic in some cell types¹⁹ although the nitrophenylethyl ester form of caged InsP_3 we have used is thought to be less toxic¹⁹. Therefore, the viability and function of guard cells loaded with the caged probes was assessed. At the light microscope level we observed: (1) no change in organelle morphology or distribution after caged probe photolysis; (2) continuation of saltatory organelle movements in the guard cells throughout the experimental period; and (3) reversal by $10\ \mu\text{M}$ fusicoccin (a fungal toxin that opens stomata²⁰) of the closure induced by Ca^{2+} release from caged InsP_3 (Fig. 3) or caged Ca^{2+} (not shown). Also, our data from caged ATP suggests that the observed closure cannot be attributed to the cage photolysis procedure.

Thus the stomatal closure we have observed in response to photolysis of caged Ca^{2+} and caged InsP_3 implicates both these

second messengers in the regulation of the events initiating this process. This study also presents preliminary evidence that InsP_3 may mobilize Ca^{2+} reserves *in vivo* in plant cells. □

Received 19 April; accepted 29 June 1990.

- De Silva, D. L. R., Hetherington, A. M. & Mansfield, T. R. *New Phytol.* **100**, 473–482 (1985).
- Schwartz, A., Ilan, A. & Grantz, D. A. *Pl. Physiol.* **87**, 583–587 (1988).
- MacRobbie, E. A. C. *Planta* **178**, 231–241 (1989).
- McAinsh, M. R., Brownlee, C. & Hetherington, A. M. *Nature* **343**, 186–188 (1990).
- Minta, A., Kao, J. P. Y. & Tsien, R. Y. *J. biol. Chem.* **264**, 8171–8178 (1989).
- Kao, J. P. Y., Harootyan, A. C. & Tsien, R. Y. *J. biol. Chem.* **264**, 8179–8184 (1989).
- Tsien, R. Y. & Zucker, R. S. *Biophys. J.* **50**, 843–853 (1986).
- Adams, S. R., Kao, J. P. Y., Grynkiewicz, G., Minta, A. & Tsien, R. Y. *J. Am. Chem. Soc.* **110**, 3212–3220 (1988).
- Walker, J. W., Somlyo, A. V., Goldman, Y. E., Somlyo, A. P. & Trentham, D. R. *Nature* **327**, 249–252 (1987).
- Cobbold, P. H. & Rink, T. J. *Biochem. J.* **248**, 313–328 (1987).
- Hedrich, R. & Neher, E. *Nature* **329**, 833–835 (1987).
- Hedrich, R. *et al. Bot. Acta* **101**, 7–13 (1988).
- Keller, B. U., Hedrich, R. & Raschke, K. *Nature* **341**, 450–453 (1989).
- Berridge, M. J. & Irvine, F. F. *Nature* **341**, 197–205 (1989).
- Canut, H., Carrasco, A., Graziana, A., Boudet, A. M. & Ranjeva, R. *FEBS Lett.* **253**, 173–177 (1989).
- Schumaker, K. S. & Sze, H. *J. biol. Chem.* **262**, 3944–3946 (1987).
- Alexandre, J., Lassalles, J. P. & Kado, R. T. *Nature* **343**, 567–570 (1990).
- Walker, J. W., Feeney, J. & Trentham, D. R. *Biochemistry* **28**, 3272–3280 (1989).
- Haugland, R. P. *Handbook of Fluorescent Probes and Research Chemicals* 131–134 (Molecular Probes Inc., Eugene, Oregon, 1989).
- De Silva, D. L. R., Cox, R. C., Hetherington, A. M. & Mansfield, T. A. *New Phytol.* **104**, 41–52 (1986).

ACKNOWLEDGEMENTS. This work was supported by the Agriculture and Food Research Council and Science and Engineering Research Council.

Bis-methionine axial ligation of haem in bacterioferritin from *Pseudomonas aeruginosa*

Myles R. Cheesman*, Andrew J. Thomson*†, Colin Greenwood‡, Geoffrey R. Moore* & Fahmi Kadir*

* School of Chemical Sciences and † School of Biological Sciences, University of East Anglia, Norwich NR4 7TJ, UK

THE iron-containing bacterioferritins^{1–6} contain the protoporphyrin IX haem group⁵. It has been established that *Escherichia coli* cytochrome *b*₁, cytochrome *b*₅₅₇ and bacterioferritin are identical⁷. The optical spectra at room temperature of the haem group show it to be predominantly low-spin in both the ferrous and ferric states⁸. The nature of the axial ligands binding the haem group to the polypeptide has, however, remained unknown. Low-spin, bis-coordinate haem centres in proteins typically have a role in rapid electron transfer as redox changes at the metal ion lead to little structural rearrangement⁹. There are only four amino acids with side-chains that have ligand field strengths sufficient to generate the low-spin state of haem, namely, histidine, lysine, methionine and cysteine. Hence there are, potentially, ten different pairs of these four ligands which could be discovered in electron transfer

† To whom correspondence should be addressed.

haemoproteins. To date only three have been established with certainty. They are bis-histidine, as in mammalian cytochrome *b₅* (ref. 10), methionine-histidine, typified by cytochrome *c* (ref. 11) and lysine-histidine, recently recognized by spectroscopic methods in cytochrome *f* (ref. 12). Here we report the electron paramagnetic resonance and near infrared magnetic circular dichroism spectra of the oxidized state of *Ps. aeruginosa* bacterioferritin which enable the axial ligands to be identified as the thioether side chains of two methionine residues, a ligation scheme not previously reported for haem in any protein.

Assignment of the axial ligands of low-spin ferric protoporphyrin IX can be made, uniquely in favourable cases, using a combination of EPR and NIR-MCD spectroscopy¹³. The spectral region between 800–3,000 nm of low-spin oxidized protohaem IX contains charge-transfer (CT) transitions involving electron excitation from the highest filled porphyrin π -orbital (a_{1u} or a_{2u}) to the hole in the Fe(III) 3d-shell, which is a combination of d_{xz} and d_{yz} orbitals¹⁴. As the energies of the d-orbitals are controlled by the ligand field strength of the haem axial ligands the wavelength of the porphyrin-to-ferric ion CT band varies as the ligand is changed. Hence the peak wavelength of the CT band can be used to identify haem axial ligands. The CT bands are most readily measured in proteins using magnetic circular dichroism (MCD) spectroscopy¹⁵ because the absorption spectrum in the region ~1,400–3,000 nm is partially obscured by the overtones arising from O–H, N–H and C–H vibrations. However, the intensity of the MCD signal depends, *inter alia*, upon the ground state magnetic properties and hence is several orders of magnitude weaker in the case of vibrational than electronic transitions¹³. The MCD spectrum therefore provides a method of locating low-energy electronic states without the possibility of confusion from vibrational bands. A wide range of haem proteins and model Fe porphyrins containing a variety of axial ligation states has been studied to determine the peak wavelengths of the CT bands characteristic of the ligation type and to enable unambiguous assignments to be achieved¹³.

The CT band shifts in a systematic way with ligand field such that an approximate additivity rule applies¹³. For example, on changing axial ligation in leghaemoglobin *a* (soybean) from histidine-imidazole to histidine-imidazolate the CT band moves from 1,600 nm to 1,360 nm, a shift of 1,100 cm^{-1} (ref. 16), and similarly a change from methionine-histidine to methionine-histidinate coordination in *Escherichia coli* cytochrome *b₅₆₂* causes the CT band to move from 1,860 nm to 1,550 nm, an energy shift of 1,080 cm^{-1} (ref. 17). We have used this additivity rule to assign the near infrared (NIR)-MCD spectrum of *Ps. aeruginosa* bacterioferritin.

The NIR-MCD spectrum of this protein is shown in Fig. 1. The electron paramagnetic resonance (EPR) spectrum is given as an inset. The MCD spectrum is temperature dependent, as expected from the paramagnetic state of the haem group. The

TABLE 1 Electronic *g*-values and peak wavelengths of CT bands of bacterioferritins and bis-thioether Fe(III) porphyrin complexes

	<i>g</i> -values	Wavelength maximum of CT band in MCD
Bacterioferritin		
<i>Ps. aeruginosa</i>	2.86, 2.32, 1.40	2,240 nm
<i>A. vinelandii</i>	2.88, 2.31, 1.46	2,270 nm
<i>E. coli</i>	2.88, 2.31, 1.46	2,270 nm*
<i>R. sphaeroides</i>	2.87, 2.32, 1.48	ND
Model compounds		
I	2.90, 2.37, 1.47	ND†
II	2.90, 2.29, 1.43	2,130 nm‡

Compound I, bis(tetrahydrothiophene) (meso-tetraphenylporphinato)-iron(III)-carborane; II, bis(tetrahydrothiophene) (octa-ethylporphinato)-iron(III)-perchlorate. ND, not determined.

* M.R.C., S. C. Andrews, P. M. Harrison, J. R. Guest, G.R.M., A.J.T., F.K. and C.G., unpublished data.

† Ref. 18.

‡ J. McKnight, C. A. Reed, M. R. C and A. J. T., unpublished data.

form of the spectrum is typically that of a porphyrin-to-ferric CT transition¹³. The wavelength of the prominent peak at 2,230 nm is unique, however, being the longest wavelength CT band yet identified in any low-spin ferrihaem. No structurally characterized protein is known with a CT band at this wavelength.

We deduce that bis-methionine coordinated protohaem IX will have a CT band in the observed wavelength region. Bis-histidine ligation gives rise to a CT band between 1,500 and 1,630 nm, depending on the protein environment, and methionine-histidine coordination leads to a CT band between 1,750–1,950 nm. Taking the mean wavelength positions of the CT bands this is a shift of 990 cm^{-1} to the red on replacing one histidine by one methionine group. Exchange of a second methionine ligand should cause a further shift to the red of ~1,000 cm^{-1} , yielding a predicted energy of the CT band of bis-methionine coordinated haem of 4,400 cm^{-1} , which is a mean wavelength of 2,270 nm. The range of wavelengths predicted are shown in Fig. 1. It can be seen that the measured wavelength of the peak falls well within the expected range. No other ligation scheme predicts the placement of the corresponding band in this region. This ligation is confirmed by the NIR-MCD spectrum of the model bis-thioether complex, that is, bis-tetrahydrothiophene-(octaethylporphinato) iron (III) perchlorate, which shows a CT band at 2,130 nm. (Table 1).

The EPR spectrum lends further support to this assignment. The inset in the figure shows the spectrum of *Ps. aeruginosa* bacterioferritin. The peaks indicated by *g*-values can be assigned to the low-spin ferric haem group. The table compares the *g*-values determined for bacterioferritin from four different bacterial sources and the *g*-values of two model bis-thioether Fe(III)

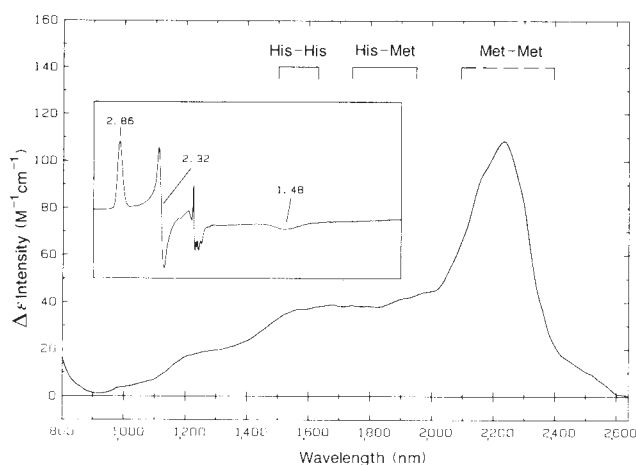


FIG. 1 The MCD spectrum of *Ps. aeruginosa* bacterioferritin at 5 Tesla and 4.2 K. $\Delta\epsilon = \epsilon_L - \epsilon_R$, where ϵ_L and ϵ_R are the molar extinction coefficients for left and right circularly polarized light, is estimated per haem group determined by double integration of the EPR spectrum. The method of measurement is given in ref. 13. The inset shows the EPR spectrum of the same sample in the region of the low-spin ferric haem signals measured between 200 and 600 mT, at 4 K, 2 mW power, and a frequency of 9.38 GHz. Haem concentration 280 μM , 200 μM HEPES buffer (pH 7.4) in D_2O with 50% (v/v) deuterated glycerol added.

porphyrin complexes. Model compound I has been shown by X-ray crystallography to have two thioether ligands coordinated to Fe(III)¹⁸. The similarity between the *g*-values of these model compounds and bacterioferritin is striking.

The evidence from the NIR-MCD spectrum and the comparison between the *g*-values of the protein and bis-thioether model compounds leaves little doubt that the haem group in *Ps. aeruginosa* bacterioferritin is ligated by the thioether side chains of two methionine residues. Table 1 indicates that this seems to be a common feature amongst bacterioferritins from a variety of sources. In no bacterioferritin has a haem content of greater than 0.5 haem per protein subunit yet been observed. For example, neither *Azotobacter vinelandii*⁴ nor *E. coli*²⁰ bacterioferritin apparently contains more than 0.5 haem per subunit and the protein from *Ps. aeruginosa* is reported to have only one haem per five subunits⁵. If 0.5 haem per subunit is the maximum number of haem groups which can bind to bacterioferritin, either the haem group is bound at the interface between a pair of subunits or only half the subunits bind haem. The identification of the haem ligands as methionine does not, at this stage, rule out the latter possibility, but we consider it most likely that the

haem group is bound by individual subunits, as is found in other multimeric haemoproteins.

Andrews *et al.*¹⁹ have reported the amino-acid sequence of *E. coli* bacterioferritin and a secondary structure analysis which indicates a high α -helical content consistent with a structure composed of bundles of four α -helices. Such a tertiary structure for bacterioferritin is consistent with its similarity to mammalian ferritins¹, shown by X-ray crystallography to consist of bundles of four α -helices^{1,21}. *E. coli* cytochrome *b*₅₆₂ (ref. 22) and *Rhodospirillum rubrum* cytochrome *c*' (ref. 23) also have a four- α -helix bundle structure. In these two proteins the haem lies between two pairs of helices so that it is located at one end of the bundle with the haem normal perpendicular to the bundle axis. On the basis of the analysis of Andrews *et al.*¹⁹, a similar location for the haem of bacterioferritin is possible, involving ligation in the *E. coli* protein by Met 1/Met 144 near the N and C termini of the polypeptide chain. Ligation by Met 31/Met 119 or Met 120 may be sterically possible. A partial sequence of the related protein from *Nitrobacter winogradskyi*²⁴ shows that Met 31 is not conserved, however, and hence cannot be a ligand, suggesting that the Met 1/Met 144 site is likely. □

Received 10 April; accepted 21 June 1990.

1. Ford, G. C. *et al.* *Phil. Trans. R. Soc. B* **304**, 551-565 (1984).
2. Lauhiere, J. P., Lescure, A. M. & Briat, J. F. *J. Biol. Chem.* **263**, 10289-10294 (1988).
3. Yaviv, J. *et al.* *Biochem. J.* **197**, 171-175 (1981).
4. Stiefel, E. I. & Watt, G. D. *Nature* **279**, 81-83 (1979).
5. Moore, G. R., Mann, S. & Bannister, J. V. *J. Inorg. Biochem.* **28**, 329-336 (1986).
6. Smith, J. M. A., Ford, G. C. & Harrison, P. M. *Biochem. Soc. Trans.* **16**, 836-838 (1988).
7. Smith, J. M. A., Quirk, A. V., Plank, R. W. H., Diffin, F. M., Ford, G. C. & Harrison, P. M. *Biochem. J.* **255**, 737-740 (1988).
8. Moore, G. R. *Biochem. J.* **227**, 341-342 (1985).
9. Moore, G. R. & Pettigrew, G. W. *Cytochromes c: Evolutionary, Structural and Physico-chemical Aspects* 370-372 (Springer, New York, Heidelberg, 1990).
10. Mathews, F. S. *Prog. Biophys. molec. Biol.* **45**, 1-56 (1985).
11. Dickerson, R. E. & Timkovich, R. *The Enzymes* 3rd edn, Vol. 11 (ed. Boyer, P. D.) 397-547 (Academic, London, 1975).
12. Rigby, S. E. J. *et al.* *Biochem. J.* **256**, 571-577 (1988).
13. Gadsby, P. M. A. & Thomson, A. J. *J. Amer. chem. Soc.* **112**, 5003-5011 (1990).
14. Smith, D. W. & Williams, R. J. P. *Struct. Bonding* **7**, 1-45 (1970).
15. Cheng, J. C., Osborne, G. A., Stephens, P. J. & Eaton, W. A. *Nature* **241**, 193-194 (1973).

16. Sievers, G., Gadsby, P. M. A., Peterson, J. & Thomson, A. J. *Biochim. biophys. Acta* **742**, 637-647 (1983).
17. Moore, G. R., Williams, R. J. P., Peterson, J., Thomson, A. J. & Mathews, F. S. *Biochim. biophys. Acta* **829**, 83-96 (1985).
18. Mashiko, T., Reed, C. A., Haller, K. J., Kastner, M. E. & Scheidt, W. R. *J. Am. chem. Soc.* **103**, 5758-5767 (1981).
19. Andrews, S. C., Smith, J. M. A., Guest, J. R. & Harrison, P. M. *Biochem. biophys. Res. Commun.* **158**, 489-496 (1989).
20. Smith, J. M. A., Ford, G. C., Harrison, P. M., Yaviv, J. & Kalb, A. J. *J. molec. Biol.* **205**, 465-467 (1989).
21. Harrison, P. M. *et al.* in *Proteins of Iron Storage and Transport* (eds Spik, G. *et al.*) 67-79 (Elsevier, Amsterdam, 1985).
22. Lederer, F., Glatigany, A., Bethge, P. H., Bellamy, H. D. & Mathews, F. S. *J. molec. Biol.* **148**, 427-448 (1981).
23. Weber, P. C., Howard, A., Xuong, N. G. H. & Salemme, F. R. *J. molec. Biol.* **153**, 399-424 (1981).
24. Kurokawa, T., Fukumori, Y. & Yamamoka, T. *Biochem. biophys. Acta* **976**, 135-139 (1989).

ACKNOWLEDGEMENTS. We thank SERC and the Wellcome Trust for support of this work and Professor P. M. Harrison for valuable discussions on the structure of bacterioferritin. The gift of a sample of *E. coli* bacterioferritin from S. C. Andrews, J. R. Guest and P. M. Harrison is gratefully acknowledged.

Implications of thermodynamics of protein folding for evolution of primary structures

E. I. Shakhnovich* & A. M. Gutin

Institute of Protein Research, Academy of Sciences of the USSR, 142292 Puschino, Moscow Region, USSR

NATURAL proteins exhibit essentially two-state thermodynamics, with one stable fold that dominates thermodynamically over a vast number of possible folds, a number that increases exponentially with the size of the protein. Here we address the question of whether this feature of proteins is a rare property selected by evolution or whether it is in fact true of a significant proportion of all possible protein sequences. Using statistical procedures developed to study spin glasses, we show that, given certain assumptions, the probability that a randomly synthesized protein chain will have a dominant fold (which is the global minimum of free energy) is a function of temperature, and that below a critical temperature the probability rapidly increases as the temperature decreases. Our results suggest that a significant proportion of all possible protein sequences could have a thermodynamically dominant fold.

We investigate a model of a protein chain in which protein folds are characterized by sets of coordinates of C $_{\alpha}$ atoms {*r*_{*i*}^{*m*}} where index *i* denotes the position of an amino-acid residue in

the sequence and *m* denotes the fold. Monomers are positioned in sites of a three-dimensional lattice to account for steric interactions. Energy of a fold *m*, say, is represented in a simple form:

$$E_m = \sum_{i,j} B_{ij} \Delta(r_i^m - r_j^m) \quad (1)$$

where *N* is the total number of monomeric units, *B*_{*ij*} is the interaction energy between monomers *i* and *j* when they are adjacent in space; this energy depends on the types of these monomers. $\Delta(r_i^m - r_j^m) = 1$ if monomers *i* and *j* are lattice neighbours and 0 otherwise.

The Boltzmann probability of a fold *p*_{*m*} is given by

$$p_m = \exp(-E_m/k_B T) / Z$$

where $Z = \sum_m \exp(-E_m/k_B T)$ is the partition function of the chain. *M* is the total number of folds; it grows exponentially with *N* so that $M = \gamma^N$ and γ is the number of conformations per monomer.

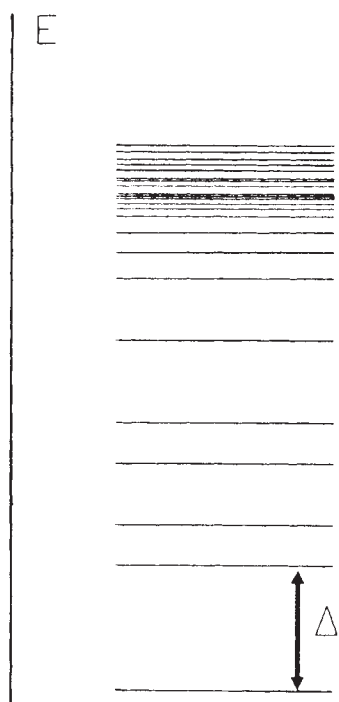
The existence of a unique structure of a protein implies that one fold *m*₀ corresponding to this structure dominates thermodynamically over all other folds so that:

$$p_{m_0} = 1 - \varepsilon \quad (2)$$

where $\varepsilon \ll 1$.

To determine the probability that a random protein folds into a unique structure, we assume the interaction energies *B*_{*ij*} to have random values. There are 20 types of amino-acid residues and, hence, 210 >> 1 types of interactions between them. Therefore we assume the probability distribution of *B*_{*ij*} to be

* Present address: Department of Chemistry, Harvard University, 12 Oxford Street, Cambridge, Massachusetts 02138, USA.



continuous and gaussian; that is:

$$P(B_{ij}) = \frac{1}{\sqrt{2\pi B^2}} \exp \left[-\frac{(B_{ij} - B_0)^2}{2B^2} \right] \quad (3)$$

where B_0 is the mean which determines overall compactness and B is the variance which corresponds to the heterogeneity of the chain.

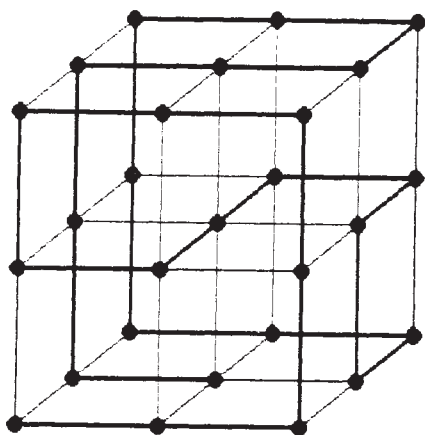


FIG. 2 The lattice and example of a compact self-avoiding walk of a chain of 27 monomers (thick line). Exhaustive enumeration by a first-depth algorithm yields all possible folds unrelated by symmetry. The total number of folds is 103,346. Two hundred realizations of sequences were generated with a distribution of contact energies between monomers B_{ij} given by equation (3). The mean B_0 is unimportant as we consider compact conformations with fixed number of contacts and this parameter gives a constant contribution to energy of each conformation. Therefore we assumed $B_0 = 0$. Standard variance B determines T_c . We estimated it according to data on inter-residue contact energies in proteins given in ref. 8: $B = 0.6 \text{ kcal m}^{-1}$ which yields $k_B T_c \approx 0.9 \text{ kcal m}^{-1}$. But the particular value of T_c is also not crucial here since we use dimensionless parameter T/T_c . Energy of each fold for each sequence was calculated according to equation (1) in which only nearest neighbours in space were assumed to interact. Boltzmann probabilities p_m ($m = 1, \dots, 103,346$) of all folds for 200 realizations of sequences (sets of B_{ij}) were determined and parameter X (equation (5)) was evaluated for each sequence at different temperatures.

FIG. 1 Typical energy spectrum of a protein. Each line corresponds to a fold. The total number of energy levels is the same as the number of folds, γ^N . The energy of each fold has a random value as we consider a random realization of a sequence. The vast majority of folds belong to the top quasicontinuous part of the spectrum, which corresponds to disordered conformations with energies lying in the range $(\bar{E} - B\sqrt{2\rho N}, \bar{E} + B\sqrt{2\rho N})$. Few levels ($\sim N$) lie in the bottom part of the spectrum; these levels correspond to folds with best-fit contacts and have energies $\sim \bar{E} - Ne$ where $e = B\sqrt{2\rho} \ln \gamma$. At high temperature the energy of a chain corresponds to the top part of the spectrum which is entropically favourable; decrease of temperature leads to the movement downwards of the spectrum and at some definite temperature, T_c , a transition takes place to the bottom discrete part of the spectrum which is entropically unfavourable but favourable energetically. The density of levels in the discrete part of the spectrum is so low that the protein cannot jump from level to level (from fold to fold) by thermal fluctuations. This is the reason for the freezing transition at $T = T_c$ when only a few lowest-lying states become available to the protein. The transition temperature T_c is universal for all sequences but the actual number of thermodynamically relevant folds at $T < T_c$ depends on specific features of the bottom part of the spectrum for a given sequence and especially on the energy gap Δ between the ground fold and the first 'excited' fold.

The thermodynamics of the chain is completely determined by the set of energies of all folds E_m , that is, by the energy spectrum (see Fig. 1). The statistical properties of this energy spectrum were obtained in ref. 1 where the microscopic model defined by equations (1) and (3) (with polymeric bonds taken into account explicitly) was investigated analytically by a replica approach. The result reads that energies of different folds can be treated as statistically independent random values with a gaussian distribution:

$$P(E_m) = \frac{1}{\sqrt{2\pi\rho NB^2}} \exp \left(-\frac{(E_m - \bar{E})^2}{2\rho NB^2} \right) \quad (4)$$

(ρ is the average number of contacts between residues) and different low-energy folds (bottom lines of the spectrum) of the same sequence are structurally different (this is the reason why their energies are independent random values).

These results demonstrate the equivalence between disordered heteropolymers and the random energy model (REM) introduced by Derrida² in the theory of spin glasses. (This equivalence had been postulated *a priori* in ref. 3).

The parameter which gives the effective number of thermo-

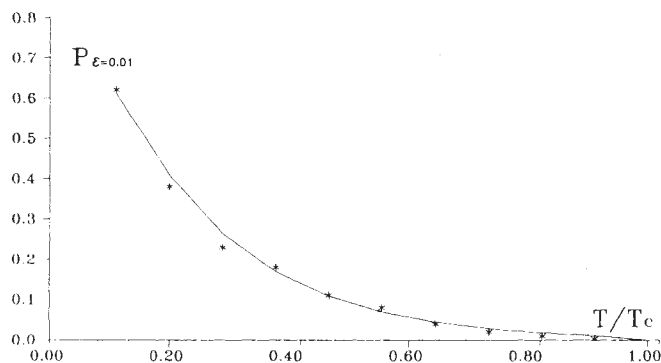


FIG. 3 Fraction of sequences that are able to fold into unique structure, that is, having Boltzmann probability of ground state $p_{m_0} > 0.99$ (or $X < 0.02$) plotted against T/T_c . The solid line denotes the analytical result of equation (6); asterisks denote the numerical results obtained from exhaustive enumeration of all conformations for 200 sequences.

dynamically relevant folds is

$$X = 1 - \sum_{m=1}^M p_m^2 \quad (5)$$

(note that $\sum_{m=1}^M p_m = 1$). When one state m_0 dominates, $p_{m_0} \approx 1$ and $X \approx 0$. In the opposite case when all microstates have equal Boltzmann probabilities, $p_m = \gamma^{-N}$ and $X = 1 - \gamma^{-N} \approx 1$.

The equivalence of random heteropolymers to the REM gives the main clue to the estimation of the probability that a random chain will fold. The probability distribution of X had been calculated for spin glasses⁴ and explicitly for the REM⁵. (For the general case of a disordered system that possesses an energy spectrum as shown in Fig. 1, with quasicontinuous top part and discrete bottom part, the same probability distribution has been derived⁶.)

Using these results, it is easy to derive the probability P_ε that for a random chain a ground fold m_0 will dominate, that is, $p_{m_0} > 1 - \varepsilon$ and, hence, to the same order in ε , $X < 2\varepsilon$:

$$P_\varepsilon = \frac{\sin(\pi X_0)}{\pi X_0} \varepsilon^{X_0} \quad (6)$$

where X_0 is X averaged over all sequences^{4,5}.

X_0 was found¹ as a function of temperature, T :

$$X_0 = \begin{cases} T/T_c & \text{if } T/T_c < 1 \\ 1 & \text{if } T/T_c \geq 1 \end{cases} \quad (7)$$

where

$$T_c = \frac{B\sqrt{\rho}}{2k_B \ln \gamma} \quad (8)$$

T_c is the temperature when the chain backbone obtains a unique conformation. It may not coincide with T_d , the temperature of denaturational transition when fixation of rotational isomers of side chains and their tight packing occurs⁷. When $T_c > T_d$, a 'molten globule' with a unique backbone conformation would exist in the temperature range $T_c > T > T_d$.

Equations (6)–(8) give the main result of this work: at high temperatures, $T > T_c$, $X_0 \equiv 1$ therefore (see equation (6)) $P_\varepsilon \equiv 0$ and there are no sequences which can fold. When the temperature decreases below T_c the fraction of sequences that are able to fold grows drastically. For example, taking $\varepsilon = 0.01$, which corresponds to 99% ground fold dominance, we obtain $P_\varepsilon \approx 0.1$ at $T = T_c/2$. This means that under this condition every tenth sequence will have one fold with 99% dominance.

This result is universal: all microscopic characteristics of folded chains such as heterogeneity B , entropy per bond $\ln \gamma$, coordination of monomers ρ , are introduced via the single parameter T_c so that the main result, equation (6) depends only on the dimensionless parameter $X_0(T/T_c)$. It should be emphasized also that the 'folding capacity' of a random chain does not depend on its length, N . The above investigation implies also that folding of a random chain could lead to the global minimum of free energy.

We tested the analytical result equation (6) by exhaustive enumeration of all self-avoiding compact conformations of a 27-monomer chain in a $3 \times 3 \times 3$ fragment of a simple cubic lattice (Fig. 2). The results (Fig. 3) match the analytical expression in equation (6).

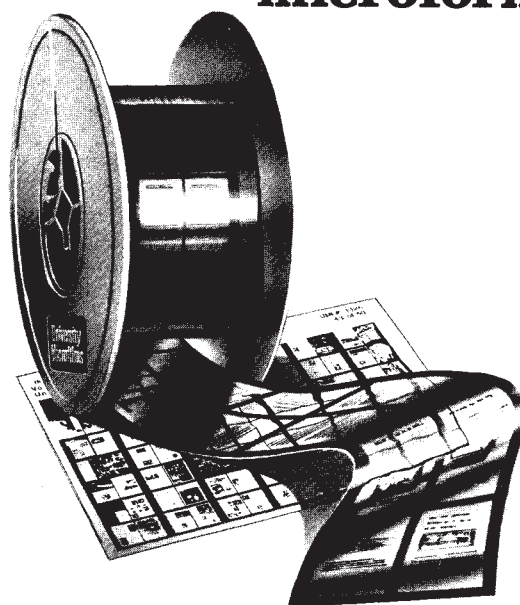
We conclude that the requirement for folding may not be very restrictive for evolutionary choice of protein sequences. $\square$

Received 30 November 1989; accepted 11 June 1990.

1. Shakhnovich, E. I. & Gutin, A. M. *Biophys. Chem.* **34**, 187–199 (1989).
2. Derrida, B. *Phys. Rev.* **B24**, 2613–2624 (1981).
3. Bryngelson, J. & Wolynes, P. G. *Proc. natn. Acad. Sci. U.S.A.* **84**, 7524–7528 (1987).
4. Mezard, M., Parisi, G., Sourlas, N., Toulouse, G. & Virasoro, M. *J. Phys. (France)* **45**, 843–854 (1984).
5. Derrida, B. & Toulouse, G. *J. Phys. (France) Lett.* **46**, L223–L225 (1985).
6. Mezard, M., Parisi, G. & Virasoro, M. *J. Phys. (France) Lett.* **46**, L217–L220 (1985).
7. Shakhnovich, E. & Finkelstein, A. *Biopolymers* **28**, 1667–1694 (1989).
8. Miyazawa, S. & Jernigan, R. *Macromolecules* **18**, 534–552 (1985).

ACKNOWLEDGEMENTS. We thank O. B. Ptitsyn for fruitful discussions. E.I.S. thanks the staff of Service de Physique Theorique C.E.N. Saclay (Paris) for hospitality during his stay in Saclay where part of this work was done.

nature is available in microform.



University Microfilms

International reproduces this publication in microform: microfiche and 16mm or 35mm film. For information about this publication or any of the more than 13,000 titles we offer, complete and mail the coupon to: University Microfilms International, 300 N. Zeeb Road, Ann Arbor, MI 48106. Call us toll-free for an immediate response: 800-521-3044. Or call collect in Michigan, Alaska and Hawaii: 313-761-4700.

☐ Please send information about these titles:

Name _____

Company/Institution _____

Address _____

City _____

State _____ Zip _____

Phone () _____

University
Microfilms
International

Biolistic transformation: microbes to mice

S.A. Johnston

A new design of the Biolistic device allows the *in-situ* transformation of animal tissue and improved efficiencies of transformation of plants and microbes.

FOR all the advances in recombinant DNA technology, the introduction of genetic material into the species or cells of interest often remains a rate-limiting step. In this regard, the Biolistic — biological and ballistic — process (Du Pont, Wilmington, Delaware) has proved to be a valuable new development. The device and procedure were originally applied to plants, and a wide variety of plant species have now been transformed¹, both transiently and stably. Recently, it was announced that the Biolistic process had been used to accomplish the long-awaited stable transformation of corn. The process has also been effective in transforming microbes, both eukaryotic² and, more recently, prokaryotic (J. Sanford, personal communication). Another first for the procedure was the stable transformation of mitochondria³. Biolistic transformation of microbial⁴ and plant⁵ chloroplasts has been reported.

Principles of operation

The original publication on biolistic transformation described several possible designs⁶. The instrument (PDS-1000) used in early experiments is now manufactured by Du Pont. The procedure involves first coating the DNA on microbeads composed of tungsten or gold. The size of the beads used ranges from 0.5 to 5 μm depending on the dimensions and penetrability of the target cells. A few microlitres of a slurry of these microprojectiles are placed on the surface of a macroprojectile, which is accelerated down a barrel by a gun-powder discharge. A stopping plate at the end of the barrel stops the macroprojectile, but a small hole in the plate allows the microprojectiles to proceed towards the target cells. The bombardment takes place in a vacuum chamber in order to lessen the impediment of air on the velocity of the microprojectiles. Several variations on this theme have been developed as instruments⁷⁻⁹ and at least one of these has been used for important advances in plant transformation¹⁰. Each of these designs involves placing the target in a partially evacuated chamber.

Animal cell transformation

An obvious question is whether the Biolistic technology can be extended to animal cells. Zelenin *et al.*¹¹ was first to

report the use of microprojectile bombardment to transform cultured mouse NIH 3T3 cells. Several other types of cultured cells have been transformed including human HT1080 and 143B (J. Sanford, personal communication), CHO, EL4 (S. McElligott, personal communication) and chick fibroblasts (D. Armaleo, personal communication). We have transformed primary cultures of both skeletal myotubes and cardiac myocytes, as well as pigmented epithelial cells. The transformation of the skeletal myotubes was highly efficient, often with greater than 10 per cent of the cells expressing the transgene product. The biolistic transformation of myotubes was 20–100 times as efficient as transformation by CaPO_4 or lipofection. Interestingly, the expression was stable for the culture life of the cells (~7 days), even though they were mitotically inactive (R.S. Williams, submitted).

Bombardment of cultured cells in the PDS-1000 leaves the centre portion of the culture heavily damaged by ballistic trauma. The helium-driven device discussed below alleviates this problem, leaving greater than 95 per cent of the cells viable. This modification has also led to significant improvements in the total expression per plate (~15-fold). These results indicate that the Biolistic transformation protocol may be useful in transforming cultured animal cells, particularly primary cells or cells recalcitrant to transformation by conventional techniques.

Helium-driven 'wand' design

One of the applications of the Biolistic technique that was anticipated was the direct transformation of tissues and organs in intact living animals¹². This goal necessitated reconfiguration of the device so that (1) it could be hand-held, (2) it would impart a high velocity to the microprojectiles, and (3) it would be non-traumatic to the tissue. The basic features of the apparatus fulfilling these demands is shown in Figure 1. The impelling force is from a shock-wave generated by bursting a membrane holding back high-pressure helium. The macroprojectile is a thin (2 mil) Kapton disc on which the DNA-coated microprojectiles have been spread. The macroprojectile is stopped by a screen in such a fashion that the target tissue is largely sealed-off from the expanding gas. The stopping screen is

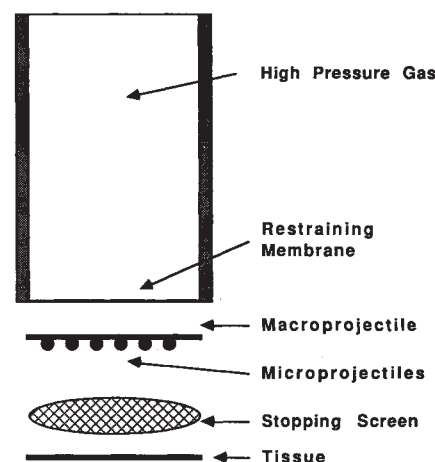


FIG. 1 Principles of operation of the helium-driven wand.

close to the target surface so that it is not necessary to create a vacuum in the flight-path of the microprojectiles, although it is possible to do so.

Using this 'wand-like' device we have successfully transformed skin, liver, spleen and intestine in living mice. The expression has been transient, using luciferase or human growth hormone as a reporter. Transgene activity has extended for up to 14 days in skin and 23 days in liver. Histological examination of the bombarded area reveals that there is relatively little damage to the tissue and that the DNA is carried into the tissue with the microprojectiles. As many as 25 per cent of the cells in the bombarded area of skin will express the transgene product (S.A.J., submitted).

The wand-like device has been fitted onto the standard PDS-1000 (see Fig. 2) producing surprising results. For every target tested the helium-driven adaption gave a 5–100-fold improvement in transformation efficiency. Yeast, bacteria, plant and mammalian cells have all responded well to the new adaption.

Future prospects

The ability to transform animal tissue *in situ* offers many opportunities. Even transient expression of transgenes could be very useful. For example, delineation of tissue-specific *cis*-acting elements could be performed directly in the authentic setting, rather than relying on transient expression assays in cell lines. Some somatic cell therapies may require only



FIG. 2 Prototype of the PDS-2000 helium-driven device.

short-term expression of some factor — for example, tissue plasminogen activator after a heart attack. Transient expression may elicit an antigenic response, making a form of genetic immunization possible.

Most somatic cell therapies, however, envision long-term or permanent expression of the transgene. The usefulness of the Biolistic technique for these applications will depend on the ability to target enough stem cells and the efficiency of integration of the introduced DNA. Whether the Biolistic device and the vectors will be adequate for these requirements remains to be seen. However, given the major contributions this technology has made to plant and microbial systems in the three years since its introduction, it seems likely that the same will be true for animal applications. □

Stephen A. Johnston is an Associate Professor at 139 Biological Sciences, Duke University, Durham, North Carolina 27706, USA. Du Pont is developing a helium-driven modification of the PDS-1000 but has at this time no specific timetable for the production of the wand-like device used in our in vivo transformation experiments. For more information, fill in reader service number 100.

- Sanford, J.C. *Physiologia Pl.* (in the press) (1990).
- Armaleo, D., Ye, G.N., Klein, T.M., Shark, K.B., Sanford, J.C. & Johnston, S.A. *Curr. Genet.* **17**, 97–103 (1990).
- Johnston, S.A., Anziano, P.Q., Shark, K., Sanford, J.C. & Butow, R.A. *Science* **240**, 1538–1541 (1988).
- Boynton, J.E. et al. *Science* **240**, 1534–1538 (1988).
- Danielli, H.J., Vivekananda, J., Nielsen, B.L., Ye, G.N., Tewari, K.K. & Sanford, J.C. *Proc. natn. Acad. Sci. U.S.A.* **87**, 88–92 (1990).
- Sanford, J.C., Klein, T.M., Wolf, E.D. & Allen, N. *J. Part. Sci. Technol.* **5**, 27–37 (1987).
- Morikawa, H., Iida, A. & Yamada, Y. *Appl. Microbiol. Biotechnol.* **31**, 320–322 (1989).
- Zumbrunn, G., Schneider, M. & Rochema, J.D. *Technique* **1**, 204–216 (1989).
- Oard, J.H., Paige, D.F., Simmonds, J.A. & Gradziel, T.M. *Pl. Physiol.* **92**, 334–339 (1990).
- Christou, P., McCabe, D.E. & Swain, W.F. *Pl. Physiol.* **87**, 671–674 (1988).
- Zelenin, A.V., Titomirov, A.V. & Kolesnikov, V.A. *FEBS Lett.* **244**, 65–67 (1989).
- Sanford, J.C. *Trends Biotechnol.* **6**, 229–302 (1988).

Washington's chemistry set

Featured at next week's American Chemical Society meeting to be held in Washington, D.C. will be a structure prediction program for NMR spectroscopists and chromatography software for IBM PCs or compatibles.

ACCORDING to Rainin Instruments, reduced solvent consumption, fast coupling and easy operation are features of the Model PS3 **automated peptide synthesizer** (*Reader Service No. 101*). Designed for unattended overnight operation if required, the PS3 can synthesize up to three separate peptides at a time, with cycle times of only 32 minutes per amino acid, says Rainin. The instrument, which sells for under \$30,000



Rainin's Model PS3 will synthesize peptides while you sleep.

(US), uses Fmoc chemistry and is designed to produce high yields without the use of hydrogen fluoride. Amino acids can be activated directly in the reagent vials, which eliminates the complex plumbing associated with conventional instruments. Reliability is assured with the PS3's inert zero dead-volume valving system. For more information on Rainin's product range, see booth 1014.

Retrovirologists may be interested in a new **simian immunodeficiency virus (SIV) p27 core antigen assay** from Coulter for the determination of SIV in plasma, serum and culture media (*Reader Service No. 102*). The assay uses a monoclonal capture EIA, similar to Coulter's HIV p24 antigen assay. Antigen that is 'captured' by anti-SIV p27 antibody after an overnight incubation is measured using the biotin/streptavidin/peroxidase detection system. The assay can be used to provide qualitative results by using positive and negative controls. Alternatively, the relative amount of antigen can be determined from a standard curve.

A collection of 67 papers on various topics in the field of **lectinology**, which were presented at the *10th International Lectin Meeting* held in Prague, Czechoslovakia, 1988, are now available in one volume from Sigma entitled *Lectins: Biology, Biochemistry, Clinical Biochemistry*, Vol. 7 (*Reader Service No. 103*).

This 462-page hard-back book includes articles by over 180 scientists in this specialized area of research.

Instrumental Ideas

Spex Industries offers the CD6 **circular dichroism spectropolarimeter system**, which automates the characterization of a compound's optical activity (*Reader Service No. 104*). As the CD6 is designed to operate with high performance down to 180 nm, it is useful for conformational studies such as the secondary structure analysis of proteins, polypeptides and polynucleotides. The instrument, which sells for under \$70,000 (US), has menu-driven software and automated control of all instrument parameters. It can be used with cylindrical, rectangular or micro-cells, without the need for adaptors. Spex also supplies a range of optional extras. Visit Spex in booth 1009.

The one-step **environmental extractor/concentrator apparatus** from Corning is designed to save time and cut costs (*Reader Service No. 105*). Made of Pyrex, the integrated system does not require sample transfer between extraction and concentration and can easily be dismantled for cleaning. The apparatus comes with inbuilt solvent recovery. Corning will be showing off its wares in booth 549.

The IR-Plan family of **FT-IR microscopes** from Spectra-Tech allow chemists and other scientists to choose the level of sampling flexibility most suited to laboratory requirements (*Reader Service No. 106*). Three models are available: IR-Plan Research, Analytical and Laboratory microscopes. Patented resolution enhancement features — redundant aperturing, targeting and sample compen-

ADVERTISEMENTS

Synthetic oligonucleotides

Custom made, purified and shipped in 3 days.
Approx. 500µg, 3.5u per base.

Oligo Express

Tel. (44) 81-420 7097.
Fax 081-420 7138

ADVERTISEMENTS

**CUSTOM-MADE & PURIFIED
DNA PEPTIDES**

\$8/base
50-100 µg
40 hour delivery

Quotes available
50-300 mg
1-3 wks delivery

GUARANTEED QUALITY & QUANTITY

V.G. CORP. 1200 Wonderland Rd. S.
Bldg #7, Unit #2
London, Ontario
Canada N6L 1A8

CALL OR FAX: (519) 652-3758

Reader Service No.29

Custom DNA

Purified and Shipped in 48 hours,
Worldwide.

100 µg, \$10.00 a base for the first 15 bases,
\$7.50 for each additional base.
Delivered.

Research Genetics

1-800-533-4363

In the United Kingdom 0-800-89-1393
In Canada 1-800-533-4363

FAX 205-536-9016

Reader Service No.26

Synthetic Oligonucleotides

Custom made Probes and Primers,
Synthesised, purified and shipped in 3 days

£4.00 per base
(All inclusive)
0.2 µM scale (approx. 500 µg)

Molecular Medicine Unit
King's College School of Medicine & Dentistry
Denmark Hill, London SE5 8RX

Tel. 071 - 326 3126 FAX 071 - 733 3877

Reader Service No.56

**Teflon^{*}
Coated
and
Bare Wire from
MEDWIRE**

Medwire provides
the highest quality
platinum, platinum alloy, silver and stainless
steel wire in both single and multistrand
configurations.

For prompt delivery or additional
information, please contact:
MEDWIRE, an affiliate of Sigmund Cohn Corp.
121 So. Columbus Ave. Mount Vernon, NY 10553
Tel: (914) 664-5300 Fax: (914) 664-5377

*Reg. TM of E.I. DuPont & Co.

Reader Service No.25

sation — ensure, says Spectra-Tech, that the sample seen and apertured is the exact sample that is represented in the infrared spectrum. Prices range from about \$19,500 to 38,000 (US). Stop by the Spectra-Tech booth, number 1220.

Computing corner

NMRgraf — the latest software tool for NMR spectroscopists available from BioDesign — is a **structure prediction program**, which uses data from NMR experiments in conjunction with computer-aided molecular design techniques, to produce and display molecular structures in 3-D (*Reader Service No. 107*). NMRgraf operates on multiple hardware platforms and typical application areas include protein and polymer research. When used with Nuclear Overhauser Effect (NOE) and J-Coupling data, NMRgraf provides a fast and accurate way of predicting low-energy structures consistent with experimental data, says BioDesign. By incorporating data from the



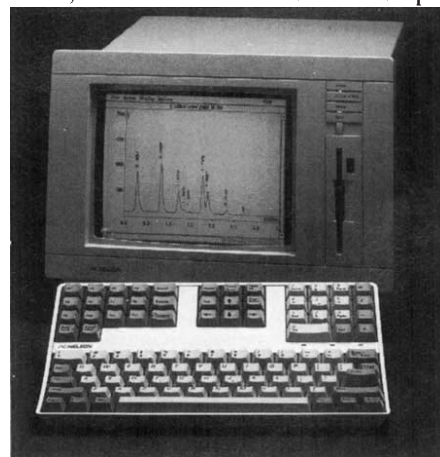
NMRgraf from BioDesign can compare predicted structures with experimental data. NMR experiment with molecular simulation, NMRgraf can be used to model most molecules, even though the structure or a related structure may not be in a database such as Brookhaven or Cambridge. NMRgraf sells for between \$40,000 and 150,000 (US) depending on CPU performance, graphics options and software modules, and can be seen in booth 814.

Version 2.22 of Lab Calc — **software for chromatographers** — is now available from Galactic Industries (*Reader Service No. 108*). Lab Calc runs on IBM PCs, PS/2s and compatibles and is a useful tool for both novice and advanced users alike. The upgraded version has a method editor, which simplifies the development of advanced chromatography methods. Lab Calc has a baseline correction feature that can subtract complex polynomial baselines from data. In addition, raw data can be adjusted for differences in run-to-run retention times and various smoothing, derivative smoothing and interpolation functions can be used to examine chromatograms. Chromatogram peaks can be zoomed in real time, peak positions, peak edges, amplitudes, areas, percent areas, concentrations or compound names can be displayed, and chromatograms can be overlaid to facilitate comparison. The software supports up to eight independent

channels for data acquisition. Data channels can be viewed together or independently or, alternatively, data collection can proceed in the background leaving the PC free to process data. See the \$1,490 (US) Lab Calc package in booth 1203.

Chemical Design's first public demonstration of its Chem-X **molecular modeling software** running on Silicon Graphics 4D Iris workstations can be seen at the ACS meeting (*Reader Service No. 109*). Workstations such as these, which are widely used by protein and polymer chemists, form a complement to the networked VaxStation and Apple computer solutions already offered by Chemical Design. Beta-testing of the product will begin in October and full release is expected by the end of the year. The annual cost of an entry-level, single-user Chem-X system for a Silicon Graphics workstation will be \$8,000 (US). See Chem-X in booth 507.

For the real-time viewing and manipulation of chromatograms and hard disk data storage, Perkin-Elmer Nelson recommends its Model 1020 **personal integrator** (*Reader Service No. 110*). Occupying just 12 inches of benchtop space, the Model 1020 with its built-in screen can be interfaced to all commercial liquid and gas chromatographs and is compatible with PCs and larger systems. Using the windows-based software, users can view one or two real-time plots, change attenuation, and examine stored or sample



Model 1020 LC/GC personal integrator.

chromatograms in detail or compared against reference standards. Chromatograms can be stored on the integrator's hard disk or saved on microdiskettes for more permanent archival storage. Perkin-Elmer will be in booth 615 at the ACS meeting.

These notes are compiled by Diane Gershon from information provided by the manufacturers. To obtain further details, use the reader service card bound inside the journal. Prices quoted are sometimes nominal, and apply only within the country indicated.